

# **Characterizing patterns in DNA sequence trace data through informatics tools**

Pedro Alberto Fernández de Mendonça

**Mestrado em Biologia Celular e Molecular**

Coimbra, September 2008







|   |             |
|---|-------------|
| <b>Acknowledgements.....</b>  | <b>IX</b>   |
| <b>Abstract.....</b>  | <b>XIII</b> |
| <b>Keywords.....</b>  | <b>XV</b>   |
| <b>1. Introduction.....</b>   | <b>1</b>    |
| 1.1 — DNA sequencing.....   | 1           |
| 1.2 — Mitochondrial DNA.....  | 2           |
| 1.3 — mtDNA and Forensic Sciences.....  | 5           |
| 1.4 — Interpreting trace sequence data.....   | 8           |
| 1.5 — Early pattern findings.....   | 9           |
| 1.6 — New pattern findings.....   | 9           |
| 1.7 — Objectives.....   | 10          |
| <b>2. Materials and Methods.....</b>  | <b>13</b>   |
| 2.1 — Reading and exporting trace sequence data.....  | 13          |
| 2.2 — Defining patterns.....  | 14          |
| 2.3 — Patterns in sequences.....  | 15          |
| 2.4 — Advanced patterns.....  | 18          |
| 2.5 — Python™ and the scripts to evaluate patterns in sequence trace data.....  | 19          |
| 2.6 — Microsoft™ Excel® as a tool to analyze the statistical significance of patterns.....  | 21          |
| 2.7 — Workflow - From the .ABI trace data file to Excel®.....   | 22          |
| 2.8 — Workflow – Usual sequence analysis.....   | 23          |
| 2.9 — Data sources.....   | 26          |
| <b>3. Results.....</b>  | <b>29</b>   |
| 3.1 — The same sample run on the same instrument with all chemistry parameters held constant produces similar sequence patterns.....                                | 29          |
| 3.2 — Different samples run on the same instrument with all chemistry parameters held constant produce similar sequence patterns.....                               | 31          |
| 3.3 — The same sample run on different instruments within the same laboratory with all chemistry parameters held constant produces similar patterns.....            | 33          |
| 3.4 — Similar samples from the same laboratory can be pooled together.....  | 34          |
| 3.5 — Samples form different laboratories with all chemistry and primer parameters held constant produce comparable, but not identical, sequence patterns.....      | 35          |
| 3.6 — Two different samples with different primers (hence different DNA sequences analyzed) but same instrument and chemistry do not exhibit the same patterns..... | 36          |
| 3.7 — Same primers, similar samples but different chemistry produce different patterns.....   | 36          |
| 3.8 — Explanation of the advanced statistical analysis.....   | 37          |
| 3.8.1 — Level of statistical significance for multiple comparisons.....   | 37          |
| 3.8.2 — Statistical significance of pattern distribution within the same laboratory.....  | 38          |
| 3.8.3 Statistical significance of pattern distribution between two laboratories.....  | 41          |
| 3.8.4 — Statistical significance of pattern distribution between two laboratories.....  | 44          |
| 3.9 — Patterns can be predicted.....  | 46          |
| <b>4. Discussion.....</b>   | <b>51</b>   |
| 4.1 — Presumptions.....   | 51          |
| 4.2 — Methods.....  | 52          |
| 4.3 — Results.....  | 56          |
| 4.4 — Sequence Biometrics.....  | 59          |
| 4.5 — Future development.....   | 60          |
| 4.6 — Final thoughts.....   | 61          |

|                            |           |
|----------------------------|-----------|
| <b>5. Conclusions.....</b> | <b>65</b> |
| <b>References.....</b>     | <b>69</b> |
| <b>Appendices.....</b>     | <b>73</b> |
| <i>Appendix I.....</i>     | 73        |
| <i>Appendix II.....</i>    | 79        |
| <i>Appendix III.....</i>   | 85        |
| <i>Appendix IV.....</i>    | 94        |
| <i>Appendix V.....</i>     | 102       |

---

## **Acknowledgements**





This work would have not been possible without the help, guidance and collaboration of many people.

Therefore, I would like to specially thank Dr. Rhonda Roby, from the University of North Texas, for putting in my hands the responsibility of solving through my computer skills the problem of evaluating and defining patterns in DNA sequence trace files, for her confidence in me, scientific guidance, deep knowledge of sequencing and unconditional help, all that without this work would be impossible. Also, I would like to thank Prof. Carol Langhauser, now retired from the University of California at Berkeley, for the amazing support on the statistical analysis of the results, her guidance and her essential teachings.

Moreover, I would like to thank Prof. Carlos Duarte, from the Universidade de Coimbra, for his willingness to provide the necessary assistance and support for the conclusion of this thesis. Now, as before, his guidance was fundamental for my work.

Besides, I would like to thank Prof. Francisco Corte-Real and Dr. Maria João Porto, from the Instituto Nacional de Medicina Legal – Delegação de Coimbra, for their collaboration and providing data for this study. A special thank you is owed to Dr. Filipa Balsa for her support, help and teachings about sequencing.

Furthermore, I would like to thank Prof. Carlos Faro and Dr. Conceição Egas from Biocant in Cantanhede, for their collaboration in providing data for this study even though, unfortunately, such data could not be used now. Nevertheless, this collaboration was fundamental for the development of the informatics tools presented in this work.

Additionally, I would like to thank Prof. Emília Duarte and Prof. António Moreno for their support in all the bureaucratic processes of this Master Degree.

Finally I would like to thank my friends for putting up with my hyped monologues about my work, my family and parents for their unconditional support and my girlfriend, Lenita, who was patient enough with me and my addiction to the computer screen, for always being there with me: I dedicate this work to you.



---

**Abstract**



Sequence trace data files produced by the Sanger sequencing method were believed to create peak height values of random height and with no added value for base calling. Our study is the first to comprehensively prove the existence of definable peak height patterns and to develop tools that allow the characterization of the frequency of such patterns for each sequence frame.

By studying hundreds of mtDNA samples sequenced in two certified forensic laboratories, in the United States of America and in Portugal, we were able to prove that peak height patterns are predictable and the same from sample to sample if the chemistry and primer combination is kept constant within the same laboratory. Moreover, the characterization of these patterns and the ability to predict their behavior for other samples led us to develop the novel concept of Sequence Biometrics. Sequence Biometrics defines the characteristics of these peak height patterns for a certain stretch of sequenced DNA, independently of the origin and sample, which is specific to the primer/chemistry combination used within the laboratory. Therefore, Sequence Biometrics is a new quality parameter for sample processing and can be used by novel expert systems in the assessment of new data.

This work provides the basic informatics tools and workflow mechanisms to build standard Sequence Biometrics tables, of several primer/chemistry combinations and to define their characteristics and boundaries of its use.



## **Keywords**

Sequence trace data

Peak height patterns

Sequence Biometrics

Expert Systems









## 1.1 — DNA sequencing

DNA is composed of four different nucleotides called adenine (A), cytosine (C), guanine (G) and thymine (T) and these four bases, combined in groups of three, provide the necessary code combination for protein translation.

Though DNA structure and the way in which bases are connected has been known since the mid 1960s, it was only in 1975 that Frederick Sanger and his colleagues revolutionized the way scientists read the genetic code by inventing the Sanger method, or the chain terminator method. Sanger and Coulson's (Sanger & Coulson, 1975) original paper described the DNA sequencing method using *Escherichia coli* DNA polymerase I and DNA polymerase from bacteriophage T4 (Englund, 1971, Englund, 1972) with different limiting nucleoside triphosphates. This method allowed scientists, for the first time, to accurately read DNA sequences base by base.

Two years later, Sanger and his co-workers described a new method for sequencing oligonucleotides via enzymatic polymerization (Sanger et al, 1977). This method, known as the dideoxynucleotide method, consists of a catalyzed enzymatic reaction that polymerizes the DNA fragments complementary to the template DNA of interest, or the unknown DNA. A  $^{32}\text{P}$ -labelled primer anneals to a specific known region on the template DNA, which provides a starting point for DNA synthesis. Catalytic polymerization of deoxynucleoside triphosphates (dNTPs) of the DNA takes place in the presence of DNA polymerases. The polymerization extends until the enzyme incorporates a modified nucleoside, known as terminator or dideoxynucleoside triphosphate (ddNTPs), into the growing chain of the synthesized DNA.

The original Sanger method was performed in four different tubes, each containing one of the four terminators. The generated fragments have the same 5'-end whereas the residue at the 3'-end are determined by the dideoxynucleotide triphosphates, or ddNTPs, used in the reaction as chain terminators since they lack the 3'-OH group required for the formation of a phosphodiester bond between two nucleotides during strand elongation. After the reactions with the four terminators, the mixture of different-sized DNA fragments is resolved by electrophoresis on a denaturing polyacrylamide gel, usually in four lanes immediately adjacent to the other. The pattern of bands is visualized and read from the autoradiography, which corresponds to the radiolabeled terminated fragments of different lengths in the synthesized strand of DNA.

Currently, fluorescent dye-labelled nucleotides are used instead of radiolabelled nucleotides. With dye-labelled dideoxy terminators, a single reaction tube can be used (Rosenthal & Charnock-Jones, 1992, Tracy & Mulcahy, 1991). The energy dyes are excited by a laser and read by a charged-couple device (CCD) optimal camera. Originally, Smith *et al.* (Smith et al, 1986) designed four different fluorescent dyes which when combined together could be electrophoresed in a single lane and read separately, since each of the dyes had its own spectral properties. This method used labels attached at the 5'-end of the primer which are attached to the ddNTP terminators. The fluorescent light is then separated by four different filters. The fluorescent dyes chosen for current use have their maximum emission spectra relatively even-spaced for better base calling. This data is then translated into human readable data; one color is associated with one base throughout the run in the capillary, creating a series of coloured peaks (given by the signal intensity of each dye) in an electropherogram that can be analyzed by software.

The Sanger method is not a high throughput method, just allowing the sequencing of merely about one thousand base pairs but it is a simple, reliable and confident method. Additionally, the results are not only rich in information but also easily readable and repeatable.

Because of its accuracy and reliability, the Sanger method has been used to sequence small genomes such as those of viruses or bacteria up to even huge genomes such as the human genome. In fact the Human Genome Project (Venter et al, 2001) based its sequencing on the Sanger method.

Furthermore, newer, sequencing procedures such as pyrosequencing, allow the sequencing of whole chromosomes in just a few hours. Yet, none of these new procedures is as informative about the DNA sequence as the Sanger method is.

## 1.2 — Mitochondrial DNA

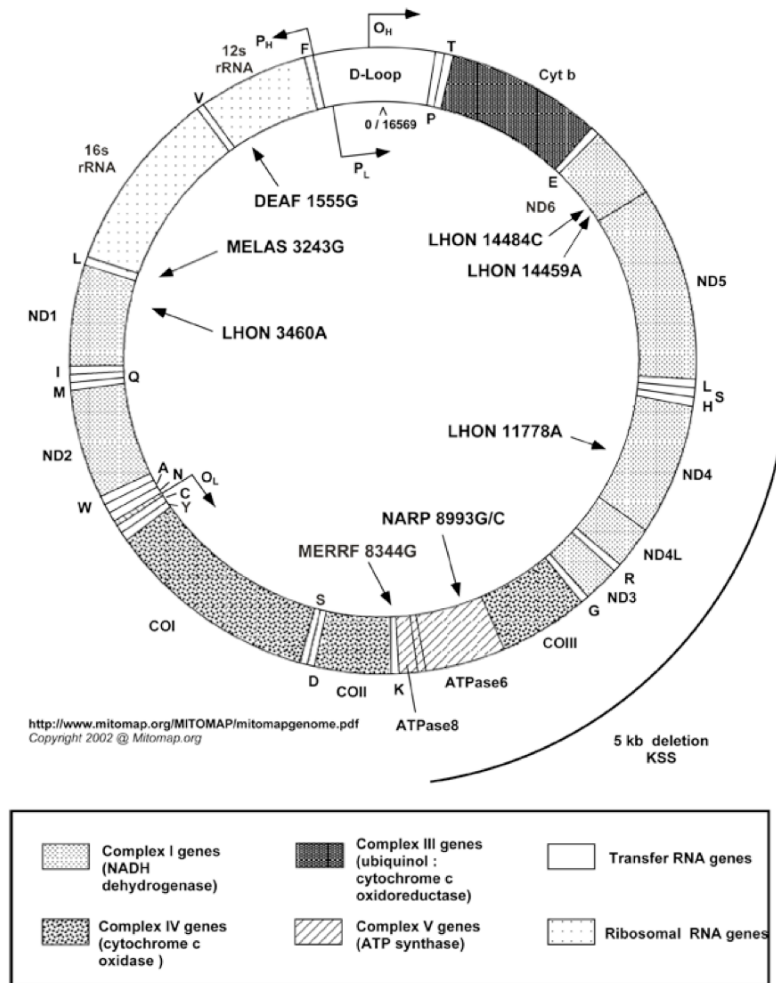
In eukaryotic cells, not all DNA is contained inside the nucleus. Organelles such as mitochondria and chloroplasts have their own reminiscent DNA that codes important proteins for their function independently of nuclear DNA.

Mitochondria are responsible for the bulk of ATP synthesis through oxidative phosphorylation and are often referred as the energy powerhouse of the cell. They are the site of

cellular respiration and capture energy generated by the breakdown of food during the oxidation of simple organic compounds (Copeland, 2002). The small number of polypeptides encoded with the mitochondrial DNA (mtDNA) genome represents only a small fraction of the total proteins necessary for mitochondrial function. Most of these proteins are encoded in the nuclear DNA genome and are subsequently exported to the mitochondria.

Mitochondria were first visualized as discrete cytoplasmic organelles in 1840 and were isolated in 1948 using zonal centrifugation techniques. They are rod-shaped organelles that are present in all nucleated eukaryotic cells that use oxygen. Mitochondria are approximately 1 to 10 micrometer ( $\mu\text{m}$ ) in length and approximately 0.5 to 1.0  $\mu\text{m}$  in diameter. Many scientists, especially evolutionary biologists, attribute the mere genetics of the mitochondrion as a primitive aerobic bacterium that was once engulfed by the ancestor of present-day eukaryotic cells (Gray, 1992, Grivell, 1997). Unlike nuclear DNA where there is only one copy from the mother and one copy from the father, most cells contain hundreds to tens of thousands discrete mitochondrion (Robin & Wong, 1988). There are exceptions, however, such as some cells containing one mitochondrion to other cells containing as many as 100,000 mitochondria (Bogenhagen & Clayton, 1974).

In the 1960s it was determined that these organelles contained their own DNA. A team of scientists at the Cambridge Research Institute completely sequenced the reported 16,569 bases of the mitochondrial genome (mtGenome) (Anderson et al, 1981). In fact, this was the first component of the human genome to be completely sequenced. DNA inside the mitochondrion is circular in structure and double-stranded. Mitochondrial DNA (mtDNA) codes for 13 polypeptides required for oxidative phosphorylation and 22 transfer RNAs and 2 ribosomal RNA subunits (see Figure 1). The heavy strand is purine-rich and the light strand is pyrimidine-rich.



**Figure 1** — The Human Mitochondrial DNA Genome. Genes encoded by the mitochondrial genome are noted. Point mutations associated with mitochondrial diseases are noted in the center of the genome. Diagram provided by MitoMap (<http://www.mitomap.org/>) (Ruiz-Pesini et al, 2007).

This closed double-stranded circular genome can be classified according to function: the coding region (about 15.5 kb of the genome) and the non-coding control region (about 1.1 kb of the genome). The control region has a regulatory function for the mitochondria and contains sequences to initiate both transcription and DNA replication of the heavy strand.

The mtGenome is not subjected to recombination during sexual transmission. The mtGenome is strictly maternally inherited; it is passed to the offspring from the oocyte (Giles et al, 1980). That is to say that progeny of both males and females inherit mtDNA from the mother (barring mutations), whereas only the daughter passes on the mtDNA to the next generation. Differences between two people indicate that they cannot share a common maternal line since it has been shown that mtDNA does not recombine in humans (Ingman et al, 2000).

The evolution of mtDNA has been studied in such detail that evolutionary biologists have determined that Mother Eve, or “mitochondrial Eve,” of all surviving mtDNA profiles lived in Africa between 140,000 and 290,000 years ago (Cann et al, 1987). The low fidelity of mtDNA polymerase and the apparent lack of mtDNA repair mechanisms have led to a higher rate of mutation in the mtGenome as compared to the nuclear genome making it an excellent marker for human evolution research. Some regions of mtGenome appear to evolve five to ten times the rate of single copy nuclear genes (Brown et al, 1979, Budowle et al, 2000). Some of the unique features of mtDNA as compared to nuclear DNA are presented in Table I.

**Table I** – Feature Comparison between Nuclear DNA and Mitochondrial DNA

| Features      | Nuclear DNA                      | Mitochondrial DNA      |
|---------------|----------------------------------|------------------------|
| Structure     | Linear genome                    | Closed circular genome |
| Size          | 3,200,000 Kb                     | 16.5 Kb                |
| Copy Number   | 1 from mother; 1 from father     | 100s to 10,000s        |
| Inherited     | 50% from mother; 50% from father | 100% from mother       |
| Ploidy        | Diploid                          | Haploid                |
| Mutation Rate | Low                              | Higher than nDNA       |
| Recombination | Yes                              | No                     |

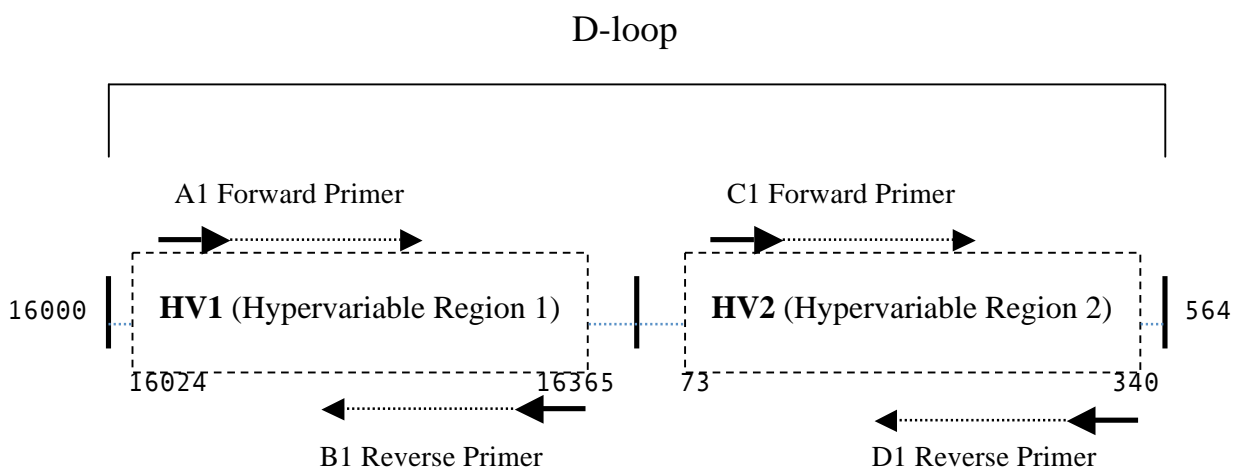
### 1.3 — mtDNA and Forensic Sciences

As described above, mtDNA has a higher mutation rate than nDNA and proceeds from the mother bloodline. Moreover, mtDNA exhibits another major characteristic: resistance to outside factors and aggressions. Being small, circular and easily condensable, mtDNA is more protected against endonucleases than nuclear DNA. Therefore, it is more resistant to degradation caused by decomposition and environmental factors.

The most interesting area of mtDNA, as far as forensic sciences are concerned, is the displacement loop, or D-loop, which is the non-coding segment of the mtGenome that maintains elements for initiation of transcription and replication but does not code for any gene products. This is the region of mtGenome that the forensic community routinely sequences for forensic casework. Since the D-loop is a non-coding segment of DNA, variability within this region is conserved which

is highly significant to the forensic scientist, since differences detected between individuals allow forensic investigators to understand that they do not belong to the same maternal line.

Many forensic laboratories have focused on sequences within the non-coding control region of the mtGenome that exhibit a higher mutation rate than the normal average rate on mtDNA, more specifically, hypervariable regions 1 and 2 (HV1 and HV2) (Figure 2). Both regions were chosen because they exhibit a large number of polymorphisms and there is a high degree of variation between individuals. These positions are ordered according to the origin of replication and numbered according to the published standard reference sequence (Anderson *et al*, 1981).



**Figure 2** — Schematic of the D-loop of the mtGenome. The D-loop is an approximate 1.1 Kb fragment of the mtGenome. In this figure, HV1 and HV2 are shown within the D-loop and their common amplification primers.

In theory, HV1 covers positions 16024 to 16365 and HV2 covers positions 73 to 340 and their sequence is given by the revised Cambridge Reference Sequence (rCRS) (Andrews *et al*, 1999) (Figure 3).



---

Hypervariable Region 1

```

15971      ttaactccac cattagcacc caaagctaag attctaattt aaactattct
16021 ctgttctttc atggggaagc agatttgggt accaccaag tattgactca cccatcaaca
16081 accgctatgt atttcgtaca ttactgccag ccacatgaa tattgtacgg taccataaat
16141 acttgaccac ctgtagtaca taaaaacca atccacatca aaacccccct cccatgctta
16201 caagcaagta cagcaatcaa cctcaacta tcacacatca actgcaactc caaagccacc
16261 cctcaccac taggatacca acaaacctac ccacccttaa cagtacatag tacataaagc
16321 catttaccgt acatagcaca ttacagtcaa atcccttctc gtccccatgg atgaccccc
16381 tcagataggg gtcccttgac caccatcctc cgtgaaatca atatcccgca caagagtgct
16441 actctcctcg ctccgggcc ataacacttg ggggtagcta aagtgaactg tatccgacat
16501 ctggttccta cttcagggtc ataaagccta aatagcccac acgttcccct taaataagac
16561 atcacgatg

```

---

Hypervariable Region 2

```

  1 gatcacaggt ctatcacct attaaccact cacgggagct ctccatgcat ttgtatattt
  61 cgtctggggg gtatgcacgc gatagcattg cgagacgctg gagccggagc accctatgtc
 121 gcagtatctg tctttgattc ctgcctcatc ctattattta tcgcacctac gttcaatatt
 181 acaggcgaac atacttacta aagtgtgtta attaattaat gctttagtagga cataataata
 241 acaattgaat gtctgcacag ccActttcca cacagacatc ataacaaaaa atttccacca
 301 aacccccctt CCCCgcttc tggccacagc acttaaacac atctctgcca aacccccaaa
 361 acaagaacc ctaaccaccg ctaaccaga tttcaaattt tatcttttgg cggtatgcac
 421 ttttaacagt ccccccaa ctaacacatt atttcccct cccactccca tactactaat
 481 ctcatcaata caacccccgc ccatcctacc cagcacacac acaccgctgc taacccata
 541 cccgaacca accaaacccc aaagacaccc cccaca

```

[http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=AC\\_000021](http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=AC_000021)

---

**Figure 3** — The Sequence of Hypervariable Regions 1 and 2. The sequence of HV1 and HV2 and their corresponding position number (Andrews *et al*, 1999).

All of these factors make mtDNA an obvious target for forensic identification of individuals. Though not individualizing *per se*, mtDNA allows the determination of a maternal bloodline, and can be easily compared and matched to individuals of that same bloodline. Besides, some mutations in mtDNA can also identify different population groups. Over the years, these sequences have been under intense population study and most of the mutations are well described for specific groups. Besides, due to the nature of forensic investigation, standard sequencing techniques have been developed that are not only highly reliable but also reproducible. Also, due to the big demand of such studies by the forensic community, these two regions (HV1 and HV2) have been sequenced over and over, using similar primers, sequencers, chemistries and protocols. This phenomenon produced a huge database of comparable sequences.

## 1.4 — Interpreting trace sequence data

In order to study differences and mutations in mtDNA, the forensic scientist must be able to read and find differences in trace sequence files.

The usual procedure of analysis for DNA sequence traces involves converting the raw data to an analyzed .AB1 file by pressing a button in the software application that came with the sequencer and then usually printing the resulting electropherogram for further analysis.

Even with the advent of computers, most forensic scientists still rely on their experience and know-how to study sequences, going through each and every base of the DNA sequence trace.

When doubts appear, scientists will go back to the sequence file to dig further into the electropherogram to clarify the sequence.

This is obviously an expensive, laborious and time consuming task that requires scientists to go back and forth in the sequence, spending a lot of time and money just to be able to study the sequence. Some software companies have been working hard to provide scientists with advanced systems: programs that can analyze trace sequence data, flag changes and orient the scientist work in a more focused and productive way.

Yet, to date, no program has been developed that is simple, reliable and a serious alternative to pen and paper for forensic analysis. This is not just a problem of software design but also mainly a problem of what quality parameters the software looks at and the response it can give to the scientist once it analyzes sequences. To date, most software packages are based on a quality assessment called Phred, which is an application capable of base calling and provide computed quality scores to those base calls defining peak signal-to-noise ratio and overall quality of for the given peak. These quality values are called Phred scores (Ewing *et al*, 1998). Though very important and informative, Phred scores alone cannot distinguish a glitch on the sequence from an interesting feature such a mutation or mixture. Also, though the overall sequence trace quality can be given, these scores say little about what interesting features that might exist.

It is then of the utmost importance for the development of better software that new quality parameters are characterized so that software can analyze sequences in a comprehensive and integrated way and also judge sequences based on a new parameter: context.

## 1.5 — Early pattern findings

Ten years ago, a revolutionary work established the first real importance of patterns in DNA sequences and suggested a new way to produce quality assessments from sequence data: by looking at the peak heights (Zakeri *et al.*, 1998).

Peak heights values, by themselves, mean very little. They are dependable on concentration and the signal produced by the dyes when passing the CCD. They vary constantly and can be different even if the same sample is run on the sequencer again. Though their absolute value is meaningless, Zakeri *et al.* were the first to propose that the way they relate to each other does, indeed, have a meaning.

This group was the first to look at a frame of three bases and describe how to measure the peak by defining the height at the higher inflexion point of the peak or the highest point of the peak. By knowing peak heights they devised a simple qualitative way of, using a relative height system, study the behavior of each frame and predict the height, within the frame, of the other two bases when given the base and height of just one of the bases in the frame.

At the time their results were promising yet mostly inconclusive. Having only studied part of the problem and very few sequences, they were not able to prove any solid relation between the height of the last peak and the rest of the peaks and the relations between chemistries.

Yet, they were the first to address the issue and, though their approach to the problem was superficial and misrouted, they did excel at defining patterns and were the first to suggest the idea of using peak height pattern information in integrated software tools to gather additional quality information on the sequenced DNA. Unfortunately their work had very little impact at the time and was cast to oblivion until recently, when new evidence suggested other ways of looking at the problem and technology evolved so that today we can output and manipulate peak height information in a much faster automated way.

## 1.6 — New pattern findings

In 2006, Rhonda K. Roby from the University of North Texas started a new approach and looked at peak height information in a novel way. If Zakeri *et al.* had just looked at the last height, Roby started looking at the entire frame and comparing the heights of all three peaks. Moreover, being a forensic geneticist, she was able to look at a bigger batch of similar, but different sample,

sequences. Though she used a very rough, by hand, study technique (much as Zakeri *et al.* work), this time, her preliminary results were not inconclusive at all. The distribution of some patterns did, in fact, appear and have statistical significance on some types of base combinations. Hence, it started to make sense that peak height patterns were not random and perhaps could be a good quality assessment tool for sequence analysis.

## 1.7 — Objectives

This work focuses on the construction of computer tools that allow the analysis of a much bigger volume of trace data and hence, a more comprehensive and exhaustive characterization of peak height patterns in trace sequence frames. The goal of this work is to define procedures and methods to look at the peak heights of DNA sequence trace data and be able to translate sequences into frames and their respective patterns. Moreover, we also want to be able to predict patterns for each frame, define what is affecting pattern formation and provide the bases for the development of improved DNA sequence analysis tools: the expert systems. The characterization of peak height pattern distributions could lead to a new important quality assessment parameter for DNA sequence trace data and for the first time be able to provide contextualized quality scores for data.

**Chapter 2** ————— **Materials and Methods**



## 2.1 — Reading and exporting trace sequence data

Every company that makes sequencers also produces the software tools to read and export the data produced during the sequencing procedure. In addition, every major sequencing company produces its own type of files. Mostly, they are closed source code, that is, proprietary and can only be read using the company software. Yet, companies usually distribute free software that can read their own files.

Applied Biosystems is, perhaps, the major company in the sequencing business. All of the data, taken from the sequencer is stored in a proprietary .AB1 file. This file includes the actual data taken from the run and also appends metadata to it such as run procedures, dates, settings, etc. Since these file types have become so popular, a great effort has been made, throughout the years, to reverse engineer the files and obtain the actual data taken from the instrument. Nowadays, many open source tools that have been developed and are able to read and export data out of .AB1 files. Recently, Applied Biosystems noticed this phenomenon and decided to share the information concerning .AB1 file format and opened it to everyone. Currently, writing and reading .AB1 files is simple and can be done reliably with most sequence analysis tools.

Some of these tools include Sequence Scanner Software, by Applied Biosystems themselves, Mutation Surveyor<sup>®</sup> by SoftGenetics LLC, 4Peaks by A. Griekspoor and Tom Groothuis (mekentosj.com) and FinchTV by Geospiza INC.

All of these tools can read .AB1 files and some can even write new information into the files. Mutation Surveyor<sup>®</sup> packs an exclusive feature built first for this study but now available to everyone: it can output the data taken from the sequencer into a simple tab-delimited text file. The file produced can then be processed using the methods described later. This unique feature is of major importance because it allows the manipulation of trace sequence data in a way that was not possible before. It is, surely, of crucial importance for this investigation.

## 2.2 — Defining patterns

As described earlier in the introduction, the existence of patterns in the heights of DNA sequence trace data is not something unknown, but it was not well described.

Yet, in order to describe patterns, one must firstly define what patterns are.

For each point where a base call is done, a single dye will be predominant within the four dyes. And the intensity signal at that point, for that dye, will be the height of the peak for that base.

For simplicity sake, frames of triplets were chosen, as they are the ideal candidates for this kind of study. Including three height values allows a descriptive, but not over descriptive analysis of the data. These three peak heights are described in Table II:

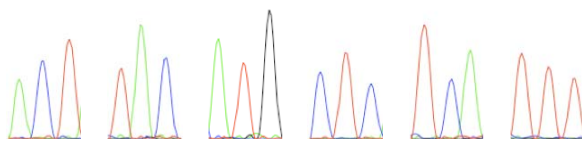
**Table II** – Height definition for Peaks

| Minimum | Medium | Maximum |
|---------|--------|---------|
| 1       | 2      | 3       |

So, it can be assumed the existence of three different heights for each base call that can be combined in six different ways, named A, B, C, D, E and F as explained in Table III:

**Table III** – Pattern definitions from heights

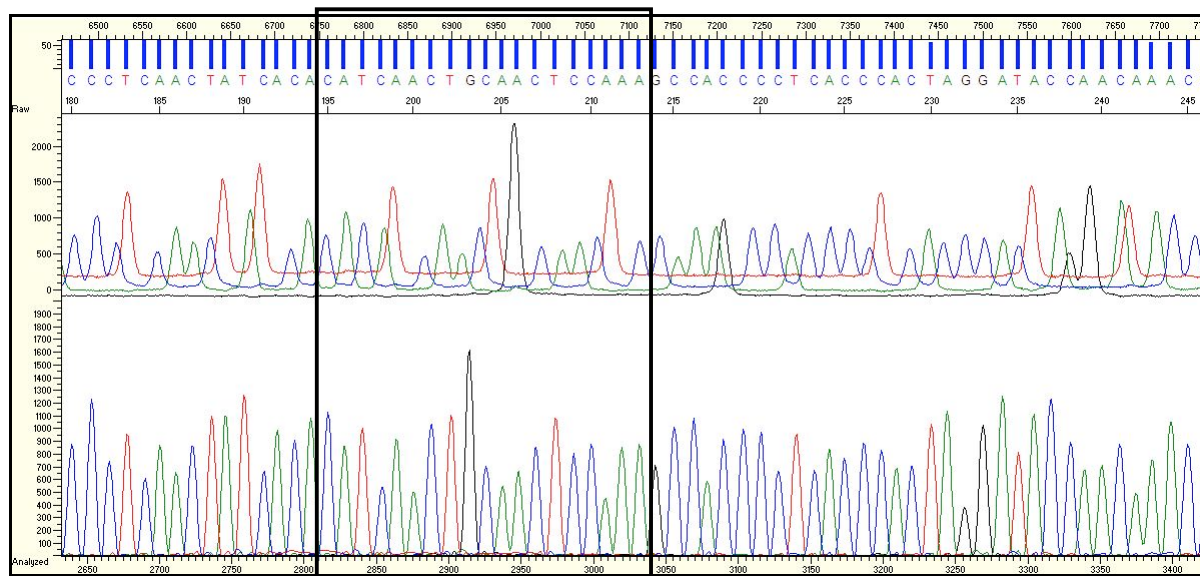
| A     | B     | C     | D     | E     | F     |
|-------|-------|-------|-------|-------|-------|
| 1 2 3 | 1 3 2 | 2 1 3 | 2 3 1 | 3 1 2 | 3 2 1 |





## 2.3 — Patterns in sequences

Now that patterns are defined, it is necessary to look at the sequences and understand how patterns are found throughout the sequence trace (Figure 4).

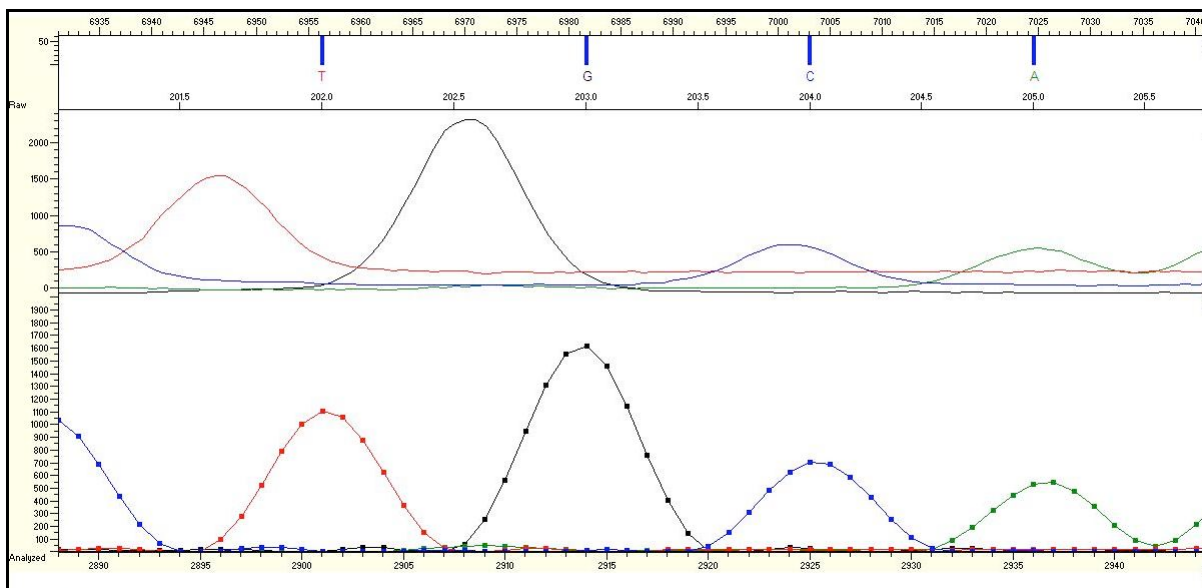


|                |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <b>Frame</b>   | CAT | ATC | TCA | CAA | AAC | ACT | CTG | TGC | GCA | CAA | AAC | ACT | CTC | TCC | CCA | CAA | AAA |
| <b>Pattern</b> | E   | D   | E   | D   | C   | A   | A   | D   | F   | E   | A   | A   | D   | E   | D   | E   | A   |

**Figure 4** — Patterns in sequences: above, an example sequence is shown with base calling, raw data and analyzed sequence trace data depicted. The box highlights one area where patterns were characterized for each frame, as it is seen in the Table of frame and patterns.

Each frame is characterized advancing a single base on the sequence at a time and describing the three peak heights for that frame in a qualitative way. It is, therefore, simple to go from the trace sequence file and find patterns for each frame. This process can be done by hand but can also be automated because traces are formed by specific data points connected with a line. These data points can be exported and, for each frame, different dye fluorescent values can be described (Figure 5)

Characterizing patterns in DNA sequence trace data through informatics tools

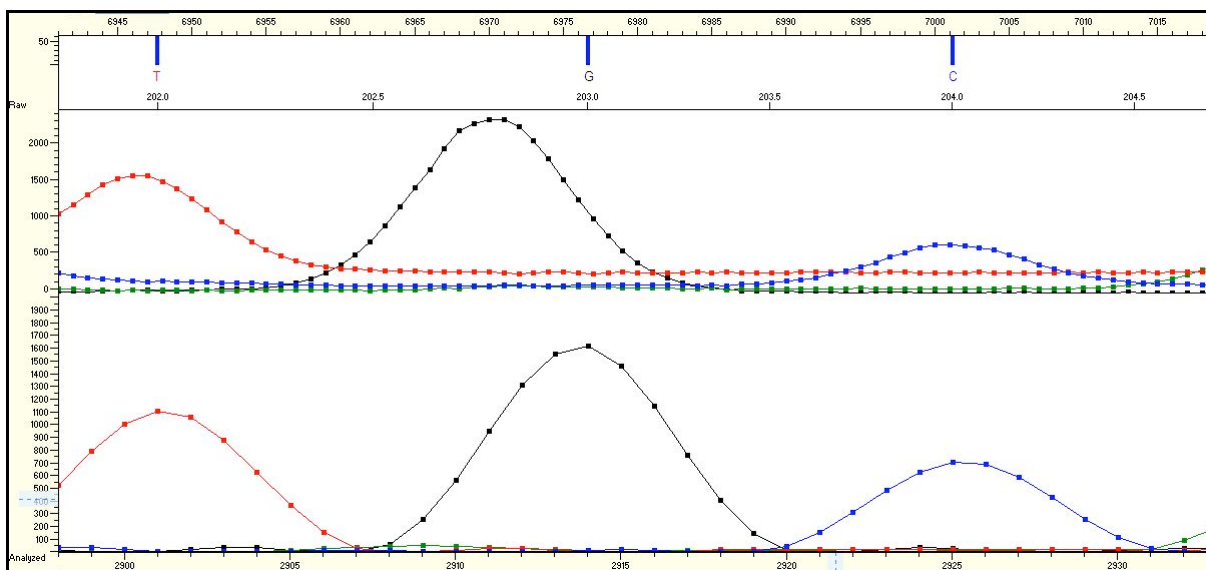


| Relative position | Green | Blue | Yellow | Red  | Base Call | Phred Score | Base Position |
|-------------------|-------|------|--------|------|-----------|-------------|---------------|
| 2004              | 7     | 1030 | 19     | 12   | C         | 59          | 200           |
| 2005              | 6     | 899  | 19     | 19   |           |             |               |
| 2006              | 6     | 631  | 17     | 29   |           |             |               |
| 2007              | 6     | 344  | 13     | 26   |           |             |               |
| 2008              | 5     | 125  | 12     | 8    |           |             |               |
| 2009              | 3     | 13   | 13     | 0    |           |             |               |
| 2010              | 0     | 0    | 17     | 3    |           |             |               |
| 2011              | 0     | 9    | 18     | 129  |           |             |               |
| 2012              | 0     | 32   | 13     | 367  |           |             |               |
| 2013              | 1     | 34   | 3      | 672  |           |             |               |
| 2014              | 1     | 21   | 0      | 950  |           |             |               |
| 2015              | 0     | 4    | 5      | 1097 | T         | 59          | 201           |
| 2016              | 0     | 0    | 23     | 1035 |           |             |               |
| 2017              | 0     | 0    | 35     | 800  |           |             |               |
| 2018              | 7     | 2    | 23     | 493  |           |             |               |
| 2019              | 20    | 6    | 3      | 220  |           |             |               |
| 2020              | 34    | 8    | 0      | 45   |           |             |               |
| 2021              | 45    | 7    | 74     | 0    |           |             |               |
| 2022              | 48    | 2    | 334    | 3    |           |             |               |
| 2023              | 42    | 0    | 735    | 18   |           |             |               |
| 2024              | 30    | 0    | 1178   | 25   |           |             |               |
| 2025              | 17    | 1    | 1515   | 14   |           |             |               |
| 2026              | 7     | 10   | 1610   | 0    | G         | 53          | 202           |
| 2027              | 4     | 15   | 1394   | 0    |           |             |               |
| 2028              | 6     | 7    | 991    | 2    |           |             |               |
| 2029              | 11    | 0    | 549    | 12   |           |             |               |
| 2030              | 14    | 0    | 200    | 19   |           |             |               |
| 2031              | 14    | 45   | 12     | 22   |           |             |               |
| 2032              | 8     | 178  | 0      | 20   |           |             |               |
| 2033              | 2     | 368  | 8      | 18   |           |             |               |
| 2034              | 2     | 560  | 29     | 18   |           |             |               |
| 2035              | 6     | 685  | 27     | 19   |           |             |               |
| 2036              | 11    | 693  | 8      | 21   | C         | 53          | 203           |
| 2037              | 11    | 569  | 0      | 23   |           |             |               |
| 2038              | 3     | 375  | 0      | 22   |           |             |               |
| 2039              | 0     | 186  | 5      | 19   |           |             |               |
| 2040              | 15    | 54   | 19     | 16   |           |             |               |
| 2041              | 80    | 3    | 27     | 15   |           |             |               |
| 2042              | 202   | 0    | 26     | 16   |           |             |               |
| 2043              | 355   | 7    | 22     | 17   |           |             |               |
| 2044              | 484   | 14   | 19     | 19   |           |             |               |
| 2045              | 538   | 9    | 17     | 21   | A         | 59          | 204           |
| 2048              | 183   | 0    | 13     | 20   |           |             |               |
| 2049              | 72    | 4    | 10     | 20   |           |             |               |
| 2050              | 72    | 5    | 7      | 21   |           |             |               |

**Figure 5** (previous page)— Data workflow: From points to peak height values. The image displays the data points for the relative positions 2004 to 2050 as seen in Sequence Scanner Software. The same points are shown in the table obtained from Mutation Surveyor®. This table contains all the height values for the analyzed points in this region, the base call for any dye, position, and quality scores for each of those base calls.

As stated before, peaks are the result of height points for each dye that are joined by a line. This study uses the data contained in each base call; yet, much more data is included in each peak. These data could be used, but decisions were made to rely on the basecaller and just analyze the height values of each dye for the points where base calls were made.

Raw data contains much more information than processed sequence data. This happens because raw data is analyzed by the basecaller which finds peaks, baselines heights, assigns bases and smoothes the peaks. This is clear in Figure 6:



**Figure 6** — Comparison of the number of points from raw to analyzed sequence trace data

All in all, the process of describing frames and calculating patterns is simple and can be easily automated. By knowing the fluorescent values of each dye at the base call it is possible to calculate the relative heights for each peak on the frame and then advance a base to find a new frame and repeat the process over and over until the end of the sequence.

This is the basic procedure for finding patterns for each frame and describing frame patterns in each sequence trace file.

## 2.4 — Advanced patterns

Sometimes, because of the resolution of the sequencer, peak heights on more than one base call, within the same frame, can be identical or very similar. In these cases, new patterns, called groups, must be defined (Table IV).

**Table IV** – Group definitions from heights

| <b>Alfa (<math>\alpha</math>)</b> | <b>Beta (<math>\beta</math>)</b> | <b>Gamma (<math>\gamma</math>)</b> | <b>Phi (<math>\phi</math>)</b> | <b>Chi (<math>\chi</math>)</b> | <b>Psi (<math>\psi</math>)</b> | <b>Null</b> |
|-----------------------------------|----------------------------------|------------------------------------|--------------------------------|--------------------------------|--------------------------------|-------------|
| 3 1.5 1.5                         | 1.5 3 1.5                        | 1.5 1.5 3                          | 1 2.5 2.5                      | 2.5 1 2.5                      | 2.5 2.5 1                      | 2 2 2       |

In this study, groups are only defined if the peak height values for two bases are identical. It could be defined a threshold of the sequencer error where patterns could be similar. Yet, for simplicity, this latter option was discarded.

The existence of these groups makes possible that, what was one type of pattern can become one type of group if the sequencer sensitivity is not enough. These combinations are explicit on Table V:

**Table V** – Possible group combinations by pattern if there are two equal minimum peaks or two equal maximum peaks

|       | <b>A</b> | <b>B</b> | <b>C</b> | <b>D</b> | <b>E</b> | <b>F</b> |
|-------|----------|----------|----------|----------|----------|----------|
| 2 Min | $\gamma$ | $\beta$  | $\gamma$ | $\beta$  | $\alpha$ | $\alpha$ |
| 2 Max | $\phi$   | $\phi$   | $\chi$   | $\psi$   | $\chi$   | $\psi$   |

But there is still another observation with the sensitivity of the sequencer: Flip-Flopping.

If two of the peak heights are equal or very similar, the same sequence run twice in the sequencer could show different values for each of those heights due to the sensitivity of the device. If this happens, then it is possible for the two peak heights to exchange places. If one was higher

than the other by just a little, it can become lower and hence, produce a different pattern. These are called Minor changes, or flip-flops, if, and only if, thresholds are not considered. Flip-flops are quite frequent in trace sequence data. Following this reasoning, Major differences occur when more than two heights in the frame change or changes are above defined thresholds. Major differences are much more significant, as they are much less possibly caused by resolution problems as Minor ones. If thresholds were considered, a minor flip-flop that would change peak heights over the defined threshold would no longer be considered a minor change but a major change. For simplicity sake, no thresholds have been considered throughout this study. Hence, group formation will only be shown when two peak height values are identical. Therefore minor and major differences will occur as described in Table VI, according to each pattern:

**Table VI** — Major/minor changes

|         | A (123)      | B (132)      | C (213)      | D (231)      | E (312)      | F (321)      |
|---------|--------------|--------------|--------------|--------------|--------------|--------------|
| A (123) | -            | Minor        | Minor        | <b>Major</b> | <b>Major</b> | <b>Major</b> |
| B (132) | Minor        | -            | <b>Major</b> | Minor        | <b>Major</b> | <b>Major</b> |
| C (213) | Minor        | <b>Major</b> | -            | <b>Major</b> | Minor        | <b>Major</b> |
| D (231) | <b>Major</b> | Minor        | <b>Major</b> | -            | <b>Major</b> | Minor        |
| E (312) | <b>Major</b> | <b>Major</b> | Minor        | <b>Major</b> | -            | Minor        |
| F (321) | <b>Major</b> | <b>Major</b> | <b>Major</b> | Minor        | Minor        | -            |

Groups, combinations and flip-flops are relevant when analyzing data for important features and its characterization also provides flexibility to the system of patterns that was described previously.

## 2.5 — Python<sup>™</sup> and the scripts to evaluate patterns in sequence trace data

Python is a very popular dynamic object-oriented programming language, based in a simple syntax and therefore very easy to learn. Python is also open source and platform independent, which makes this particular language widely available from desktops to mobile devices and web servers. Moreover, given its outstanding capabilities in handling text strings, Python has become a very

popular language for bioinformatics. All in all, Python is a great programming language for building applications capable of processing the text data used in this work.

Three scripts were written to process data out of Mutation Surveyor<sup>®</sup>:

**“Sanger”** – The pattern finder script. It recalculates base calling, builds frames and finds the patterns in each of those frames. This is the core tool. Since it is programmed in Python, its easily modifiable according to specific data changes.

**“Cleaner”** – An optional script that goes through the “Sanger” output file and finds repeating mononucleotide frames, choosing only the first. This cleaning process is important because it allows for the elimination of non-independent observations.

**“Reporter”** – This is a counting and organizing script. Reporter counts each occurrence of frames and patterns by scanning and counting the files produced by either “Cleaner” or “Sanger”. Also, “Reporter” writes its counts to a text file that can be easily imported into Microsoft<sup>™</sup> Excel<sup>®</sup> for further processing.

Special nomenclature – Some scripts have several versions and therefore may be named differently. “Cleaner\_NI” finds only non-independent mononucleotide repeats. “Reporter\_CD” is a version of “Reporter” capable of counting the number of cleaned frames after the “Cleaner” processing.

All of these scripts are run in a Terminal window by calling on the Python Interpreter. Moreover, they allow the user to specify the file to process or ask the script to process all the text files in a folder and also allow the user to choose the location and name of the output file.

Other scripts were built to help formatting the results. They are either called the comparison scripts that can compare between two sequences, frame by frame, pattern by pattern or the formatter scripts that take data from the comparison and format it so that is easily readable.

---

## 2.6 — Microsoft<sup>™</sup> Excel<sup>®</sup> as a tool to analyze the statistical significance of patterns

In order to analyze the statistical significance of patterns, special Excel<sup>®</sup> spreadsheets were developed.

Microsoft<sup>™</sup> Excel<sup>®</sup> is a simple, widespread spreadsheet environment and, even though it is not an application built to deal with sophisticated statistical analysis, it manages a simple Chi Square Goodness of Fit test well. Moreover spreadsheets in Excel<sup>®</sup> are easy to program, format and deliver or export to other applications.

Two types of spreadsheets were built to study the significance of peak height patterns.

The “1x6” spreadsheet, which performs a Chi Square Goodness of Fit test for a single data sample for each frame and analyzes the distribution of patterns, deeming them as random or not, and the “2x6” spreadsheet, which also performs a Chi Square Goodness of Fit test for a pair of samples for each frame, calculating the significance of pattern distribution of each frame, comparing both distributions for each frame and deeming them as uniform or not.

These spreadsheets are able to process and format data in a way that the results from the scripts are understandable. Also, these spreadsheets are able to calculate the statistical significance of each count for each frame, automatically, and produce results. Yet, this automated process is not error-free and all datasets have been reviewed upon import into Excel<sup>®</sup>.

After reviewing, when necessary, further statistical testing was performed to the data using the Kolmogorov-Smirnov goodness of fit test.

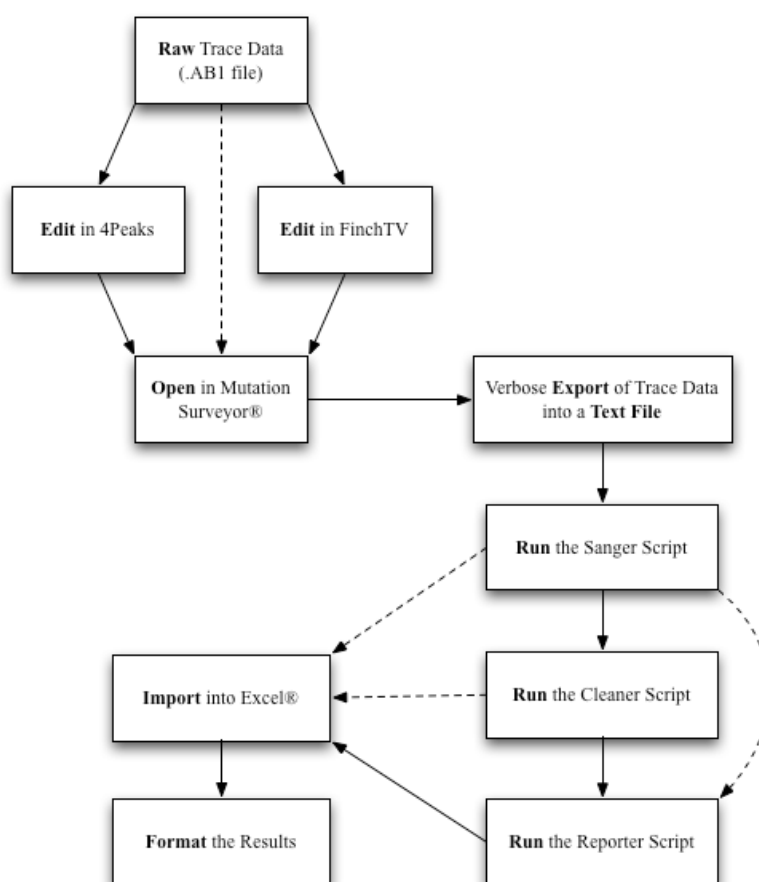
The Kolmogorov-Smirnov test is a nonparametric test of equality of one-dimensional probability distributions that is used to compare a sample with a reference probability distribution (one-sample K-S test), or to compare two samples (two-sample K-S test). The Kolmogorov-Smirnov test statistic quantifies a distance between the empirical distribution function of the sample and the cumulative distribution function of the reference, or between the empirical distribution functions of two samples. This test allows comparing our distribution of patterns with an empirical expected distribution (random in our case) and is mostly useful when there are few observations. Alas, it is not a very powerful test but it allows for the validation of the Chi Square test results

calculated with too few observations (especially in cases of distributions that show few observations all gathered around 2-3 pattern types).

The biggest drawback of the Kolmogorov–Smirnov test is that is not viable to program it into a simple Excel<sup>®</sup> spreadsheet. Yet, the Excel<sup>®</sup> spreadsheets are capable of handling all the statistic calculations that were done, since the Kolmogorov–Smirnov tests are only used on specific occasions to validate the Chi Square Goodness of Fit tests.

## 2.7 — Workflow - From the .AB1 trace data file to Excel<sup>®</sup>

Now that every step of the procedure was described by itself, from the definition of patterns or how frames are built to the statistical analysis of the results, the workflow used for processing the data will be described. The workflow used throughout this study for each batch of data is as follows (Figure 7):

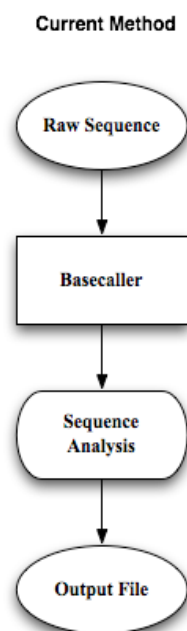




**Figure 7** (previous page)— Data workflow: from the AB1 Trace File to Excel<sup>®</sup>. Files are opened in either 4Peaks or FinchTV for editing; if not, they are opened directly in Mutation Surveyor<sup>®</sup>. Then, data are exported into a text file to be processed in the Sanger Script either individually or as a batch. Data are then processed all the way into the Reporter Script and then imported into Excel<sup>®</sup>. Dashed arrows indicate possible, yet not very common variations of the workflow.

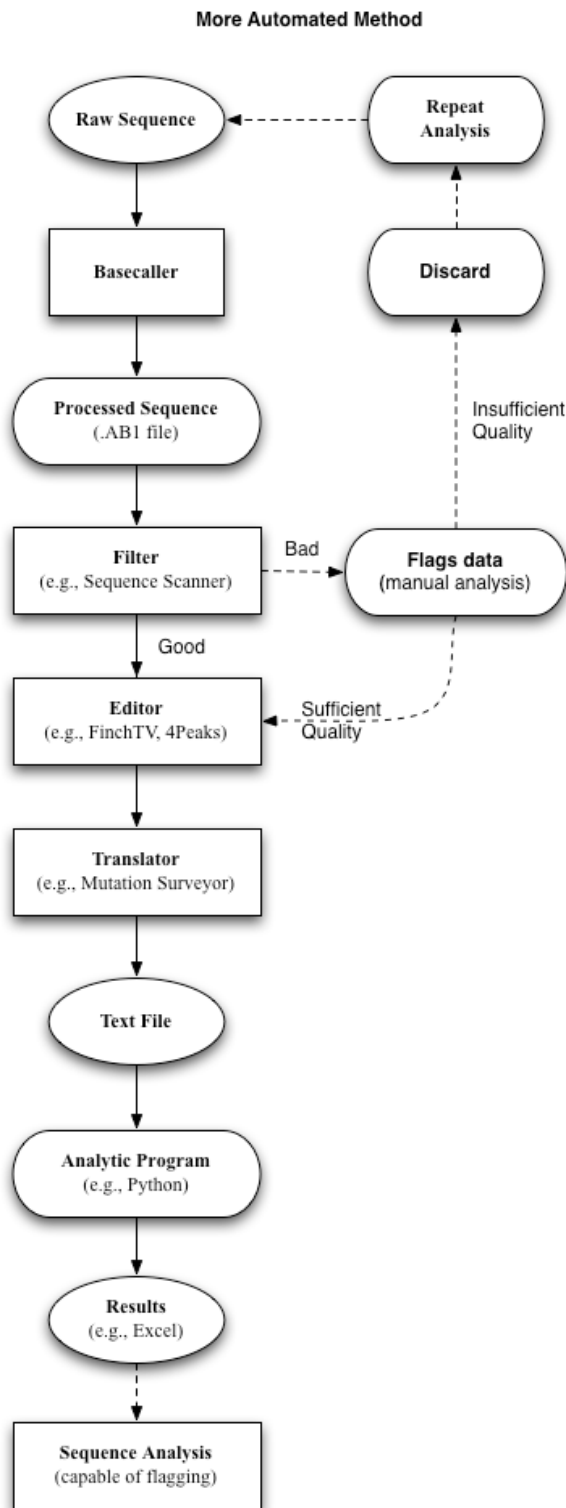
## 2.8 — Workflow – Usual sequence analysis

Most researchers follow a simple procedure to analyze DNA sequence traces. This procedure is depicted in Figure 8:



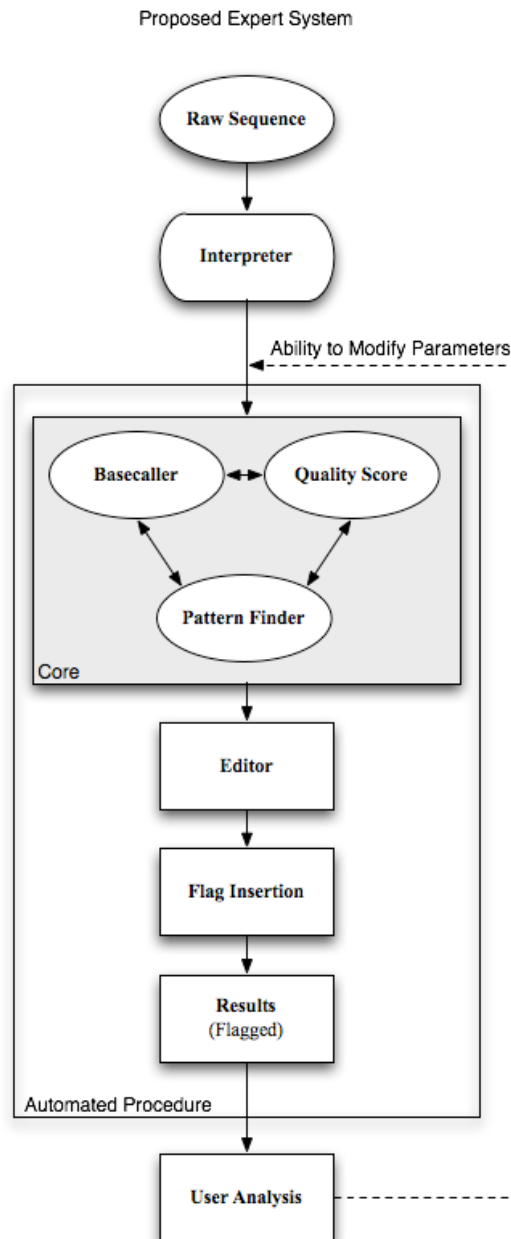
**Figure 8** — Typical workflow: the raw sequence is processed within a software package and sequence analysis is exported to the output file, which is usually printed. Users don't intervene during the process.

The methodology followed, throughout this study is quite different as it was described before. This study proposes a better solution for the workflow. The workflow used is based on the best features of some software tools combined together to empower the analysis of trace sequence data in a more comprehensive way (Figure 9).



**Figure 9** — Our workflow is based on a more complex analysis of sequence trace data. Data is processed, filtered, translated and analyzed using different software tools; however, we must intervene on every step of the process.

Looking at the earlier procedure and the way we were able to take existing tools and use bits and parts of them in order to provide a better sequence analysis, we can now propose a new way to look at and analyze sequence data: an expert system. The backbone of such system is shown in Figure 10.



**Figure 10** —Proposed expert system. An expert system is an automated software tool that can look at sequence data and, without user intervention, flag problems in the sequence. Data is fed to the system, processed and the output is ready, with all quality parameters assured, for interpretation by the analyst.

This expert system should be able to read sequence data, call bases, provide quality scores for those bases and find patterns within frames. Then, gathering these three bits of information, patterns, base call and quality, it would be able to edit and trim bad areas and/or flag bad bases, error points or mismatches. All in all, it would offer the user a digested view of the sequence rather than the simple, unprocessed results currently scientists are forced to analyze. An expert system such as this would, therefore, allow a much quicker analysis of large sequence datasets and also a more reliable and consistent way of gathering quality information from the data.

## **2.9 — Data sources**

Throughout our study we used data from routine mtDNA samples that had been processed for forensic purposes only, using similar methodologies and processing protocols: either BigDye<sup>®</sup> Terminator v.1.1 Cycle Sequencing Kits (Applied Biosystems, Foster City, CA, USA) or dRhodamine Terminator Cycle Sequencing Kits (Applied Biosystems). The sequences were read in similar ABI PRISM<sup>®</sup> 3100 and 3130*xl* Genetic Analyzers in certified high quality facilities. The data used comes from two different laboratories, the University of North Texas Health Science Center (Laboratory 1 or A) and the Serviço de Genética Forense from the Instituto Nacional de Medicina Legal (INML), Delegação Centro (Laboratory 2 or B). All of the data used is of classified nature, and stripped down of any information that is not relevant for this investigation.

Also, there was the opportunity to work with data from the sequencing facility at BioCant. Yet, for technical reasons that will be explained later, this data was not analyzed.



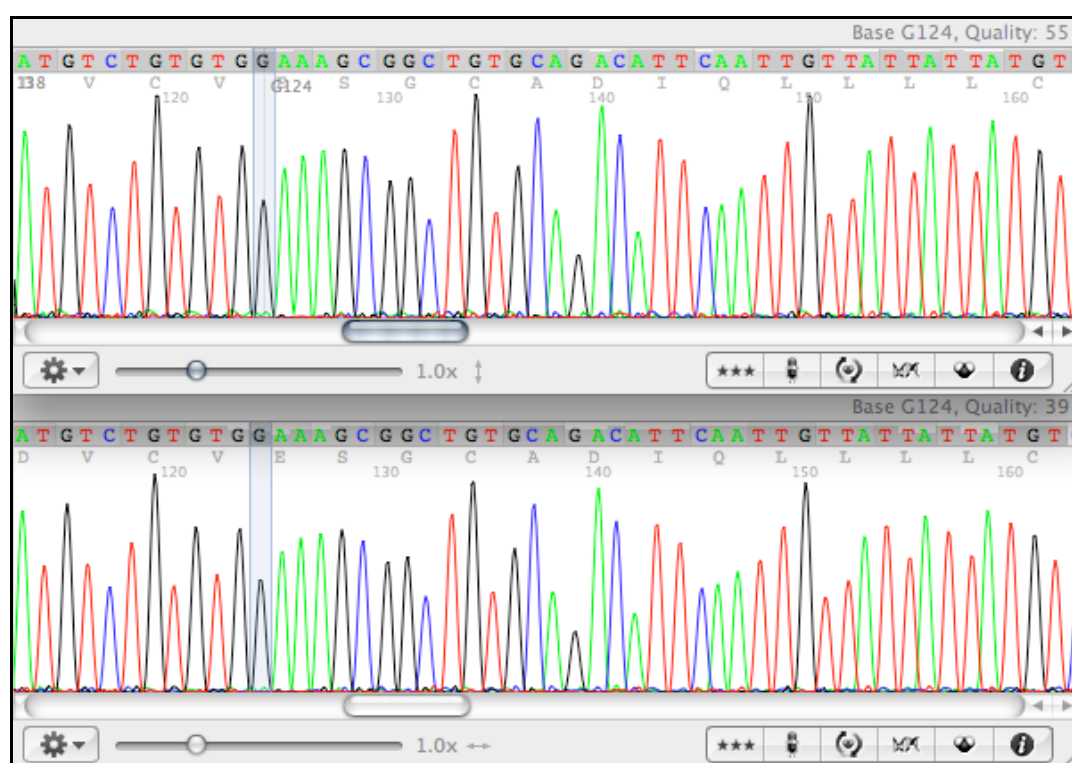
**Preliminary note about the results:**

In this study, results are always shown as a summary of the processed data. For each section, examples of the data used can be found on the indicated annex. This option has been determined due to the unusual length of the data itself, which would force a trimming of most of its contents if it were to be shown, or else a very long document. Yet, every example, though lengthy, has been chosen carefully to provide you, the reader, with a clear view of every step taken in the elaboration of the results. The data used throughout the study is available electronically on the CD that you can find on the back of this book, organized in folders by sections. Information about system requirements and copyright notice is also printed on the back of this book.

### 3.1 — The same sample run on the same instrument with all chemistry parameters held constant produces similar sequence patterns

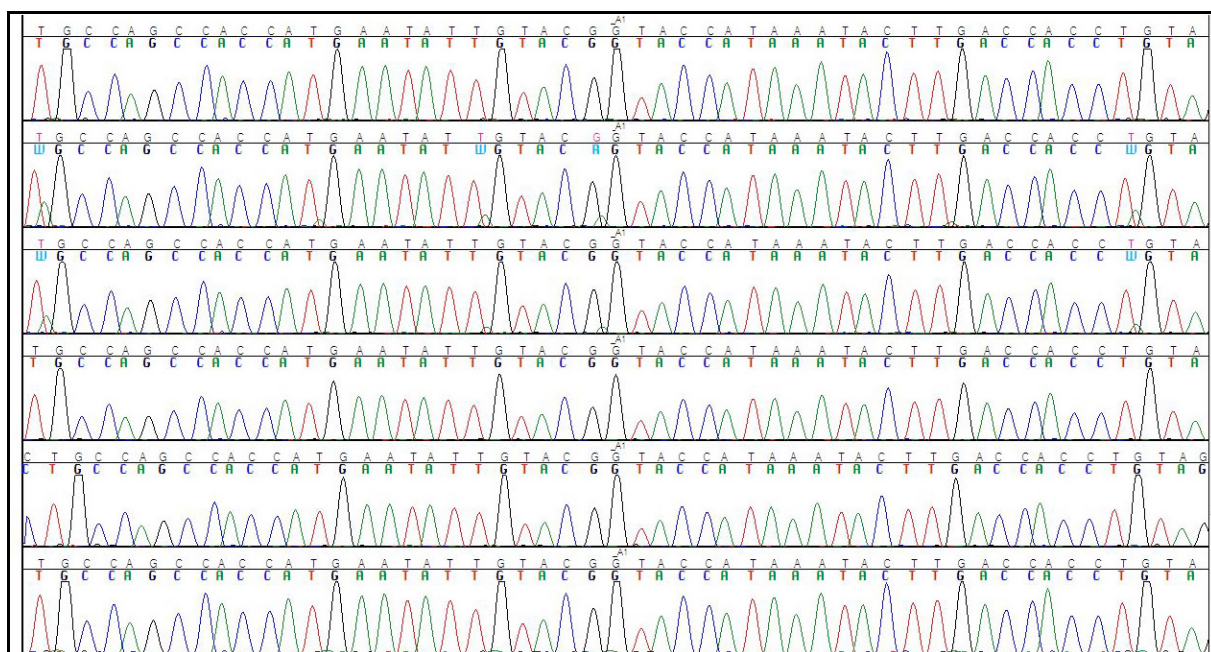
In order to prove the existence and meaning of patterns, the first question that must be answered is whether peak height patterns within each frame are kept constant if the same sequence is run several times in the same instrument. If so, this will mean that patterns are not a totally random event.

To do so, a sample of one individual was taken and distributed into two wells on the same plate and then the sequencing procedure was run using the usual laboratory routine. Figure 11 shows the two electropherograms of the same sample run twice on the same instrument:



**Figure 11** – Two electropherograms depicting the same sample run twice on the same instrument using BigDye 1.1 chemistry and POP-6. The sample is of mitochondrial origin and was sequenced using a D1 primer.

Figure 12 shows six electropherograms of the same sample run constantly over and over on the same instrument:



**Figure 12** – Six electropherograms depicting the same sample run on the same instrument using BigDye 1.1 chemistry and POP-6. The sample is of mitochondrial origin and was sequenced using an A1 primer.

As seen in both Figures 11 and 12, it is clear that peak height patterns are kept constant on multiple runs for the same sample.

In order to mathematically analyze this observation, the same sample was run twice in the same instrument (divided into two wells) and then the two sequences were, aligned, run through processing scripts and compared. Then the same was done for other primer combinations. Please refer to Table A in Appendix I for an example of the results. Table VII summarizes the results of all these comparisons.

**Table VII** – Difference comparison between the same samples run twice on the same instrument.

| Comparison | Total Frames | Diff. Frames | Dif. Patterns | Same Frame, Dif. Pattern | Major differences |
|------------|--------------|--------------|---------------|--------------------------|-------------------|
| <b>A1</b>  | 380          | 3            | 34            | 31                       | 0                 |
| <b>B1</b>  | 250          | 0            | 14            | 14                       | 0                 |
| <b>C1</b>  | 357          | 0            | 12            | 10                       | 0                 |
| <b>D1</b>  | 362          | 0            | 15            | 13                       | 0                 |

Notice that the first ten frames of each sequence were discarded because of noise and trimming artifacts and that major differences are counted if, and only if, a minor one does not



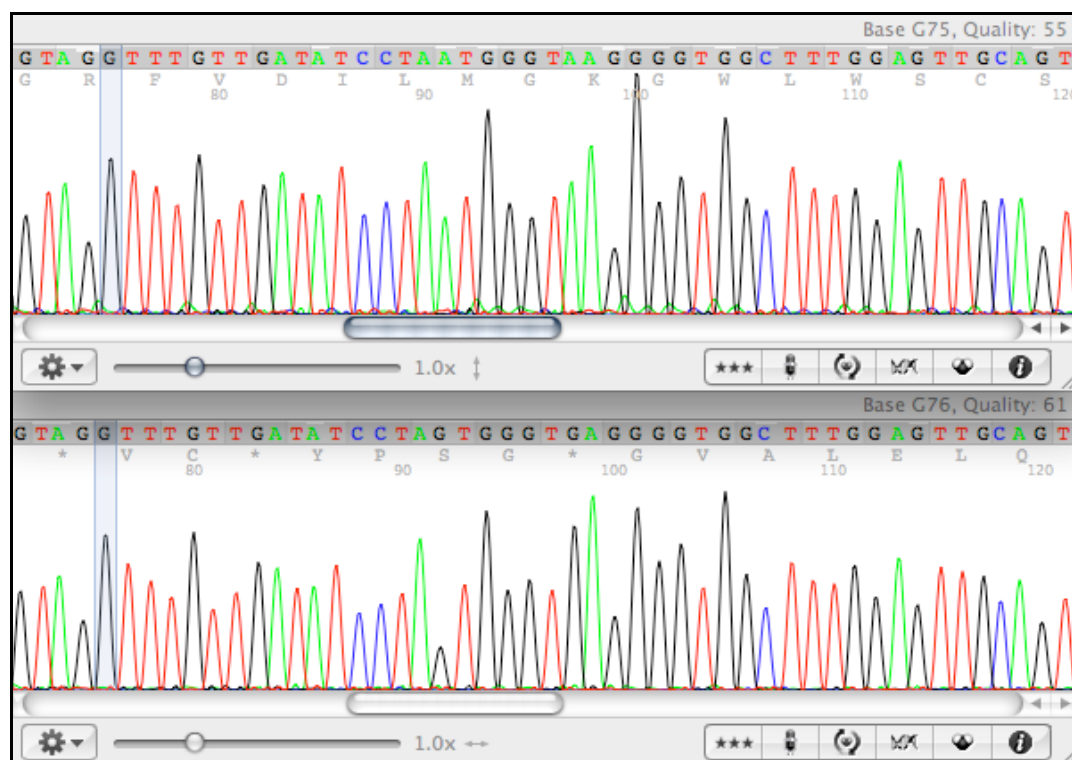
precede them because otherwise it would not be an independent event. Of relevance, notice the small number of pattern differences (<10%) and the almost absent number of different frames.

### 3.2 — Different samples run on the same instrument with all chemistry parameters held constant produce similar sequence patterns

The last section showed that patterns are kept constant when the same sample is run, using the same parameters, more than once on the same instrument.

Now, the second question to answer is whether peak height patterns are kept constant if samples are different but the parameters are kept the same. Hence, to know if two different samples behave in a similar way and show similar patterns if the parameters are kept constant and, if that is true, know if those results can be pooled together.

To do so, two different mtDNA samples belonging to two different individuals were taken and, keeping all the primer and chemistry parameters constant, the sequencing procedure was run on the same instrument. Figure 13 shows two of those samples run in the same instrument.



**Figure 13** (previous page)– Two electropherograms depicting two different samples on the same instrument using BigDye 1.1 chemistry and POP-6. The samples are of mitochondrial origin and were sequenced using a B1 primer.

As seen in Figure 13, is clear that peak height patterns are kept mostly constant from sample to sample. Bases may change, peak heights may also change as so do patterns, but the same frame will show very similar pattern behavior from sequence trace to sequence trace.

In order to mathematically analyze this, sequence traces from each of two different samples were aligned, run through-processing scripts and compared. Then, the same process was performed for other primer combinations. Refer to Table B in Appendix I for an example of the results. Table VIII summarizes the results of these comparisons.

**Table VIII** – Difference comparison between two different samples run on the same instrument.

| <b>Comparison</b> | Total Frames | Diff. Frames | Dif. Patterns | Same Frame, Dif. Pattern | Major differences |
|-------------------|--------------|--------------|---------------|--------------------------|-------------------|
| <b>A1</b>         | 148          | 8            | 35            | 27                       | 1                 |
| <b>A2</b>         | 81           | 8            | 26            | 14                       | 2                 |
| <b>B1</b>         | 220          | 18           | 66            | 47                       | 1                 |
| <b>C1</b>         | 194*         | 9            | 47            | 34                       | 1                 |
| <b>C2</b>         | 122*         | 3            | 36            | 29                       | 1                 |

\* Usable number of frames (before or after sequence mismatch). All values are calculated over this number.

Again, notice that the first ten frames of each sequence were discarded because of noise and trimming artifacts and that major differences are counted if, and only if, a minor one does not precede them because otherwise it would not be an independent event.

It is clear that, though differences do occur, most patterns are kept constant from one sample to the other. Moreover, very few major differences occur (no more than two significant differences in any comparison).

This indicates that, though samples are different, similar sequences will produce similar patterns. Moreover it indicates that, because they are similar, different traces from different samples analyzed will show enough similarities so that they can be pooled together and provide the same results. This will be demonstrated later on.

### 3.3 — The same sample run on different instruments within the same laboratory with all chemistry parameters held constant produces similar patterns

Another preliminary question to answer is whether pattern behavior within frames is dependant on the chemistry and parameters used or the instrument itself.

In order to answer that, the same sample was taken and, keeping all the chemistry parameters constant, run through the sequencing procedure on two different instruments, but within the same laboratory.

These two sequence traces were, aligned, run through the processing scripts and compared. Please refer to Table C in Appendix I for an example of the results. Table IX summarizes the results of these comparisons.

**Table IX** – Difference comparison between the same sample run on two different instruments.

| Comparison | Total Frames | Diff. Frames (minus cleaned) | Dif. Patterns | Same Frames, Dif. Pattern | Major differences |
|------------|--------------|------------------------------|---------------|---------------------------|-------------------|
| <b>A1</b>  | 369*         | 0                            | 25            | 25                        | 0                 |
| <b>B1</b>  | 332*         | 0                            | 33            | 33                        | 0                 |
| <b>C1</b>  | 89*          | 0                            | 20            | 20                        | 0                 |
| <b>C2</b>  | 94*          | 0                            | 8             | 8                         | 0                 |
| <b>D1</b>  | 327*         | 0                            | 26            | 24                        | 1                 |
| <b>D2</b>  | 145*         | 0                            | 14            | 14                        | 0                 |

\* Usable number of frames (before or after sequence mismatch). All values are calculated over this number.

As usual, notice that the first ten frames of each sequence were discarded because of noise and trimming artifacts and that major differences are counted if, and only if, a minor one does not precede them because otherwise it would not be an independent event.

These results are very similar to those in section 3.1. Hence, the same sample run in two different instruments produces patterns in a similar way as if that sample were to be run twice in just one instrument. All in all, just one case of a significant major difference was found in one of the comparisons.

### 3.4 — Similar samples from the same laboratory can be pooled together

The results from the last sections suggest that patterns are kept constant within each frame, in the same laboratory, independently from the sample used if the chemistry and primer used are kept constant.

In order to prove that mathematically, batches of A1, B1, C1 and D1 sequences, using BigDye 1.1 chemistry, were pooled together and analyzed statistically for two different laboratories.

Since the process involves pooling together several sequences, comparisons had to be performed by using a simple Chi Square Goodness of Fit test. Refer to Table A and B in Appendix II for an example of the results. Table X summarizes these results.

**Table X** – Analysis of the predictability of patterns within frames.

| <b>Lab/Primer</b> | Predictable<br>Frames | Too little<br>information | More than 4<br>blank patterns | Low Chi<br>Square | Random<br>patterns |
|-------------------|-----------------------|---------------------------|-------------------------------|-------------------|--------------------|
| <b>Lab A A1</b>   | 55                    | 2                         | 5                             | 2                 | 0                  |
| <b>Lab B A1</b>   | 52                    | 5                         | 6                             | 1                 | 0                  |
| <b>Lab A B1</b>   | 54                    | 3                         | 5                             | 2                 | 0                  |
| <b>Lab B B1</b>   | 54                    | 4                         | 5                             | 1                 | 0                  |
| <b>Lab A C1</b>   | 63                    | 0                         | 1                             | 0                 | 0                  |
| <b>Lab B C1</b>   | 56                    | 0                         | 7                             | 1                 | 0                  |
| <b>Lab A D1</b>   | 54                    | 0                         | 9                             | 1                 | 0                  |
| <b>Lab B D1</b>   | 60                    | 0                         | 3                             | 0                 | 1                  |

Predictable frames are all the frames for which the p-value of the distribution is higher than 0,00078125 and the Chi Square value is higher than 15,085, for a confidence of 95%. Values are only discarded for frames with less than 12 observations or low Chi Square value.

Notice that only one case, with 23 observations, produced a result that can be considered as a random distribution of patterns. There are 448 total predictable frames (87.5%) and 41 (8%) of all frame pattern observations were grouped in less than two types of patterns. Only about 5% of the frames had too little information, a low Chi Square value or showed a random distribution of patterns. It is then safe to assume that, given these results, peak height patterns within the same laboratory, for the same chemistry and primer are kept constant.

This is an important milestone. Not only peak height patterns are not random but also they can be grouped according to the primer and chemistry used for the same laboratory because they are the same. These results are consistent and corroborate the results found in earlier sections.

### 3.5 — Samples from different laboratories with all chemistry and primer parameters held constant produce comparable, but not identical, sequence patterns

In section 3.1 and 3.3 it was already addressed that the same sample run in different instruments will produce similar patterns. It is also known, from section 3.2 and 3.4, that sequences from different samples can be pooled together into batches if, and only if, they are produced in the same laboratory using the same chemistry and keeping all the parameters constant. All in all, the distribution of patterns is kept constant within a laboratory in these conditions.

Now, the question is whether similar sequences will produce similar results in two different laboratories or not. It is important to know if sequence peak height patterns can be exported from one laboratory to the other if the parameters are kept constant or if these patterns are a characteristic of the laboratory that produced them.

To answer this question a batch of A1, B1, C1 and D1 sequences were pooled together from two different laboratories and compared. Refer in Appendix III to Table A and B for an example of the results. Table XI summarizes the comparisons.

**Table XI** – Comparison of frame patterns for the same chemistry in two different laboratories.

| <b>Primer</b> | Comparable | Not comparable | Too little information | More than 6 blank cells |
|---------------|------------|----------------|------------------------|-------------------------|
| <b>A1</b>     | 30         | 19             | 2                      | 13                      |
| <b>B1</b>     | 38         | 13             | 4                      | 9                       |
| <b>C1</b>     | 32         | 28             | 0                      | 4                       |
| <b>D1</b>     | 35         | 16             | 0                      | 13                      |

These results show that pattern formation from Laboratory A to B for any given primer is not as highly similar as it is from sample to sample within the laboratory or from instrument to instrument within the laboratory as it was previously shown. About a third of the time, the two laboratories exhibit frame patterns that are significantly different. Yet, the frame patterns formed in Laboratory A and Laboratory B are globally uniform and these results could be merged. Of notice, unmatched frames do not exhibit, sometimes, the same types of patterns. Sometimes, though pattern distribution for the frame is different, one of the pattern types is always missing, which might be a symptom of flip-flopping.

**3.6 — Two different samples with different primers (hence different DNA sequences analyzed) but same instrument and chemistry do not exhibit the same patterns**

To this point it has been made clear that, using the same chemistry and primers, the same region of the DNA, when sequenced, will produce similar patterns independently from the instrument used, but patterns may change from laboratory to laboratory. The next step is to understand if different regions of DNA will produce similar patterns within the same laboratory, that is, if patterns are determined only by the frame chosen or, if not, dependant also on the entire DNA sequence that is being sequenced.

To do so, one could take individual samples and compare them, but it is cleverer to look at batches of A1 sequences and compare them to B1, C1 and D1 sequences. Refer in Appendix IV to Table A and B for an example of the results. Table XII summarizes the results of these comparisons.

**Table XII** – Comparison of frame patterns for the same chemistry but two different primers.

| <b>Primers</b> | Comparable | Not comparable | Too little information | More than 6 blank cells |
|----------------|------------|----------------|------------------------|-------------------------|
| <b>A1 – C1</b> | 2          | 52             | 0                      | 10                      |
| <b>B1 – D1</b> | 4          | 51             | 0                      | 9                       |
| <b>A1 – B1</b> | 3          | 52             | 0                      | 9                       |

Since A1 sequences produce different types of patterns, for each frame than B1, C1 or D1 sequences, it becomes clear that different primers will produce different types of patterns for each frame, even within the same laboratory. Hence, peak height patterns in sequence data, though predictable, are also dependent on the DNA sequence.

**3.7 — Same primers, similar samples but different chemistry produce different patterns**

At this point it is known that the distribution of patterns are reporducible within frames if the laboratory, chemistry and primers are the same. Now, the remaining question is whether different chemistries produce different patterns or not.

To answer this question a batch of similar samples, processed in the same laboratory using different dye chemistries (dRhodamine or BigBye<sup>®</sup>) was taken for A1, B1, C1 and D1 primers.

These results were then compared for significance.

Refer in Appendix V to Table A and B for the results. Table XIII summarizes the results of these comparisons.

**Table XIII** – Comparison of frame patterns for the same primer but different chemistries.

| <b>Primer</b> | Comparable | Not comparable | Too little information | More than 6 blank cells |
|---------------|------------|----------------|------------------------|-------------------------|
| <b>A1</b>     | 5          | 54             | 0                      | 5                       |
| <b>B1</b>     | 3          | 57             | 0                      | 4                       |
| <b>C1</b>     | 1          | 63             | 0                      | 0                       |
| <b>D1</b>     | 5          | 59             | 0                      | 0                       |

\* Usable number of frames (before or after sequence mismatch). All values are calculated over this number.

The results show that the chemistry used is also a major player in the formation of patterns for each frame. Different dye chemistries radically alter the distribution of patterns within each frame. Hence, the DNA area sequenced will influence the distribution of patterns, but also the dye chemistry used to obtain such sequences will alter the distribution of patterns for each frame. This means that even though if samples are sequenced using the same primer, changing only the chemistry within the same DNA sequence area will change the distribution of patterns.

### 3.8 — Explanation of the advanced statistical analysis

#### 3.8.1 — Level of statistical significance for multiple comparisons

For section 3.4 to 3.7 (Appendices II through V), 64 frames were evaluated for statistical significance. At a significance level of 5% for a single test, the Type I Error is  $\alpha = 0.05$ . As an example a significance level of 5% for each of 20 different independent tests, the Type I Error is very high. The probability of at least one error is 1 minus the probability of making no Type I Errors, or  $1 - P$  [no Type I Errors], and can be found using the binomial distribution when each test is independent of all other tests. The binomial distribution where  $n$  equals the number of tests (*e.g.*,  $n = 20$ ) and  $p$  equals the probability of a Type I Error (*e.g.*,  $p = 0.05$ ) is defined as:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{where the mean} = \mu = np \text{ and the variance} = \sigma^2 = np(1-p)$$

$$P(X = 0) = \binom{20}{0} (0.05)^0 (1-0.05)^{20} = (.95)^{20} = 0.358$$

So,  $P(X > 0) = 1 - P(X = 0) = 1 - 0.358 = 0.642$  for 20 independent tests.

And the expected number of Type I Errors =  $np = 20(0.05) = 1.0$  for 20 independent tests. The probability of making no Type I Errors in 20 independent tests is 0.358; therefore, the probability of making at least one Type I Error is  $1 - 0.358 = 0.642$ , or 64.2%. For 40 independent tests, the probability of at least one Type I Error is  $P(X > 0) = 1 - .95^{40} = 1 - 0.129 = 0.871$  and the expected number of Type I Errors is 2.0.

Therefore, for 64 independent tests, *i.e.*, 64 sequence frames, the probability of at least one Type I Error is  $P(X > 0) = 1 - .95^{64} = 1 - 0.0375 = 0.962$  and the expected number of Type I Errors is 3.2. At a significance level of 5%, it would not be surprising to find one or more significant tests by chance alone. In other words, the Type I Error associated with all 64 comparisons is no longer 5% but a higher value, namely 96.2%. The more tests that are performed the more likely it is that "significant results" will be found by chance when, in fact, no significant difference exists.

When there are "multiple comparisons," an adjustment should be made to keep the overall Type I Error to a minimum. One way to do this is to divide the  $\alpha$  level by the number of independent tests (*i.e.*, 64 sequence frames; therefore,  $\alpha/64 = .05/64 = 0.00078$ ). This yields a Type I Error of  $P(X > 0) = 1 - (1 - .00078)^{64} = 1 - (0.9992)^{64} = 1 - 0.951 = 0.049$ . This adjustment keeps the overall level of significance at approximately 0.05 for 64 independent tests. However, there is a corresponding loss of power,  $1 - \beta$ , which occurs whenever the critical point is changed in a direction away from the null hypothesis,  $H_0$ . The adjusted critical point is determined by  $\alpha/n$  making the rejection region smaller.

### 3.8.2 — Statistical significance of pattern distribution within the same laboratory

In section 3.4 it was studied data from two laboratories that were independently compiled and processed with our tools to establish if patterns were distributed uniformly within each frame.



---

Each of the four primers, A1, B1, C1, and D1, were evaluated to characterize the different patterns for each frame for the two different laboratories. The results from each laboratory were obtained using the same sequencing procedures. The null hypothesis is that each of the six patterns is equally likely. The p-value is the probability of observing a Chi Square greater than or equal to that observed under the null hypothesis of a uniform distribution with each cell having a relative frequency of 1/6 with 5 degrees of freedom. There are two tables in Appendix II. For Appendix II, Table A is the automated program previously described showing the Chi Square results and corresponding p-values for Primer A1 only. Table B presents the p-values for all of the primers for Laboratory A followed by a table for Laboratory B. The question being asked is:

**“Within a frame, are the frequencies of the patterns distributed uniformly?”**

**“Are the patterns equally distributed?”**

**“Are the patterns equally likely?”**

In symbols, that is:

**$H_0: P(A) = P(B) = P(C) = P(D) = P(E) = P(F)$**

Or,

**“The patterns are equally likely.”**

Since there are 6 patterns:

**$P(A) = P(B) = P(C) = P(D) = P(E) = P(F) = 1/6$**

**$H_1: H_0$  is false**

Some patterns are more likely than others within a given frame.

The test statistic used is the Chi Square Goodness of Fit Test. The Chi Square Goodness of Fit Test is used to test how well the data fit the null hypothesis that  $P(A) = P(B) = P(C) = P(D) = P(E) = P(F) = 1/6$ .

$$X^2 = \sum_{i=1}^n \frac{(O_i - E_i)^2}{E_i} = \sum_{i=1}^6 \frac{\left(O_i - \frac{N}{6}\right)^2}{\frac{N}{6}} \text{ where } E_i = \frac{N}{6} \text{ for all expecteds}$$

The degrees of freedom, df, used in this test are determined by:

$$df = (\text{number of cells}) - (\text{number of independent parameters estimated}) - (\text{number of restrictions})$$

$$df = 6 - 0 - 1 = 5$$

The level of significance at  $\alpha = 0.05$  overall with 64 independent tests is  $0.05/64 = 0.00078125$  and the critical point with 5 df for Chi Square is 21.8322.  $H_0$  is rejected if the calculated Chi Square with 5 degrees of freedom gives  $X^2 > 21.8322$  with  $p < 0.00078125$ . Table XIV is an example of a Chi Square Goodness of Fit statistic test, which rejects the null hypothesis for frame AAA; the distribution of patterns is not equally likely. In summary, the null hypothesis for both Laboratory A and Laboratory B combined is rejected for 488 of 495 frames.

**Table XIV** — An example of a 1 x 6 Chi Square Goodness of Fit test rejected. The frame, AAA, yields very different frequencies between several patterns. Patterns A, C, and E are similar; however, B and F are much lower and D has no observations. The  $H_0$  is rejected because  $X^2 = 28.28$  which is greater than 21.8322 with  $p = 0.000032$  which is less than 0.00078125. Hence, patterns are not distributed uniformly within a frame of AAA.

| Pattern | A     | B     | C     | D     | E     | F     | TOTALS     |
|---------|-------|-------|-------|-------|-------|-------|------------|
| AAA     | 20    | 9     | 20    | 0     | 19    | 7     | 75         |
| Exp.    | 12.50 | 12.50 | 12.50 | 12.50 | 12.50 | 12.50 |            |
| $X^2$   | 4.50  | 0.98  | 4.50  | 12.50 | 3.38  | 2.42  | 28.28      |
|         |       |       |       |       |       |       | p-Value:   |
|         |       |       |       |       |       |       | 3.2086E-05 |

Table XV is an example of a Chi Square Goodness of Fit statistic test, which does not reject the null hypothesis for frame GCA; the distribution of patterns may be equally likely. In summary, the null hypothesis for both Laboratory A and Laboratory B combined is not rejected for seven of the 495 frames. For both laboratories, a total of nine frames are not applicable (N/A) since none or too few observations were seen for those frames.

**Table XV** — An example of a 1 x 6 Chi Square Goodness of Fit test not rejected. The frame, GCA, yields different frequencies between the patterns but they do not differ much from the expected distribution of:  $N/6 = 50/6=8.33$  under  $H_0$ . The  $H_0$  is not rejected since  $X^2 = 8.08$  which is less than 21.8322 with  $p = 0.15188$  which is greater than 0.00078125. Hence, patterns may be distributed uniformly within a frame of GCA.

| Pattern        | A    | B    | C    | D    | E    | F    | TOTALS |            |
|----------------|------|------|------|------|------|------|--------|------------|
| GCA            | 9    | 13   | 2    | 9    | 7    | 10   | 50     |            |
| Exp.           | 8.33 | 8.33 | 8.33 | 8.33 | 8.33 | 8.33 |        | p-Value:   |
| X <sup>2</sup> | 0.05 | 2.61 | 4.81 | 0.05 | 0.21 | 0.33 | 8.08   | 0.15188144 |

For those frames with expected values less than that required in more than 25% of the cells, a One Sample Kolmogorov test was performed. The tables are not presented here; the p-values calculated with the One Sample Kolmogorov test statistic were generally consistent with those of the Chi Square Goodness of Fit test. A total of 19 of the 493 frames were calculated with the One Sample Kolmogorov test. Table XVI represents an example of the test performed to an ACC frame.

**Table XVI** — Example of One Sample Kolmogorov test. When there are 30 or fewer total observations, which would yield an expectation less than 5.00 ( $30/6 = 5.00$ ), the One Sample Kolmogorov Test is performed. Although this test statistic is less powerful than the Chi Square, the results were generally consistent with the Chi Square Goodness of Fit test. CRF is the cumulative relative frequency; EXP CRF is the expected CRF which is always: 1/6, 2/6, 3/6, 4/6, 5/6, and 6/6 based on the  $H_0$ : Patterns are equally likely,  $P=1/6$ . |DIFF| is the absolute difference between CRF and EXP CRF. The  $H_0$  is rejected because One Sample Kolmogorov = 0.611 which is greater than 0.371 with  $p \ll 0.01$  and is consistent with rejecting  $H_0$ . Hence, the distribution of patterns is not uniform within frame ACC.

| Pattern   | A     | B     | C     | D     | E     | F     | TOTAL     |  |
|---|-------|-------|-------|-------|-------|-------|-----------|--|
| ACC   | 0.00  | 17.00 | 0.00  | 0.00  | 0.00  | 1.00  | 18.00     | Compared to<br>X <sup>2</sup> p-value<br>1.59481E-15 |
| CRF   | 0.000 | 0.944 | 0.944 | 0.944 | 0.944 | 1.000 |           |  |
| EXP CRF   | 0.167 | 0.333 | 0.500 | 0.667 | 0.833 | 1.000 |           |  |
| DIFF  | 0.167 | 0.611 | 0.444 | 0.278 | 0.111 | 0.000 |           |  |
| Critical point at 0.01 with n = 18 is 0.371<br>(limit of table) |       |       |       |       | KOLM= | 0.611 | p << 0.01 |  |

### 3.8.3 — Statistical significance of pattern distribution between two laboratories

The data from the two laboratories compiled and processed using our tools in Appendix II were subjected to another test for Appendix III. A 2 x 6 Chi Square analysis was performed to compare the results from Laboratory A to Laboratory B with the four primers. The null hypothesis is that the distribution of patterns for the two laboratories is the same. The distribution of patterns within each of the 64 possible frames for Laboratory A was compared to Laboratory B.

There are two tables in Appendix III. Table A is the output of the automated program previously described showing the Chi Square results and corresponding p-values for the two laboratories for Primer A1 only. Table B presents the summary of the p-values for all of the primers and frames comparing distribution for Laboratory A to Laboratory B.

The question being asked is:

**“Within a frame, do the two laboratories produce the same distribution of patterns?”**

**“Are the probabilities of the patterns the same for the two laboratories?”**

**“Is the distribution of patterns independent of the laboratory?”**

In symbols, that is:

**$H_0$ :  $P(A | \text{Lab 1}) = P(A | \text{Lab 2}) = P(A)$  and  $P(B | \text{Lab 1}) = P(B | \text{Lab 2}) = P(B)$  and  $P(C | \text{Lab 1}) = P(C | \text{Lab 2}) = P(C)$  and  $P(D | \text{Lab 1}) = P(D | \text{Lab 2}) = P(D)$  and  $P(E | \text{Lab 1}) = P(E | \text{Lab 2}) = P(E)$  and  $P(F | \text{Lab 1}) = P(F | \text{Lab 2}) = P(F)$**

That is, for  $H_1$ ,

**$H_1$ :  $H_0$  is false**

**“At least one of the patterns is different between the two laboratories.”**

**“The distribution of the patterns is not independent of laboratory.”**

Table XVII is an example of a 2 x 6 Chi Square Goodness of Fit statistic test, which rejects the null hypothesis for frame ACC; the two laboratories do not produce the same distribution of patterns for Frame ACC using Primer C1.

**Table XVII** — An example of a 2 x 6 Chi Square Goodness of Fit test for independence rejected (Appendix III, C1, same chemistry). The frame ACC yields very different relative frequencies within several patterns as can be seen by comparing the observed to the expected and Chi Square contributions between the two laboratories. Patterns A and F are similar while C, D, and E are different and make major contributions to the overall Chi Square. The  $H_0$  is rejected because the calculated  $X^2 = 56.87$  which is much greater than 21.8322 with  $p = 5.38378E-11$  which is much less than 0.00078125. Hence, the distribution of patterns is not independent between the two laboratories.

| Pattern        | A      | B      | C      | D     | E      | F      | TOTALS |                                       |
|----------------|--------|--------|--------|-------|--------|--------|--------|---------------------------------------|
| ACC            | 242.00 | 89.00  | 144.00 | 35.00 | 98.00  | 105.00 | 713.00 |                                       |
| ACC            | 77.00  | 41.00  | 13.00  | 28.00 | 8.00   | 34.00  | 201.00 |                                       |
| TOTAL          | 319.00 | 130.00 | 157.00 | 63.00 | 106.00 | 139.00 | 914.00 |                                       |
| Exp.           | 248.85 | 101.41 | 122.47 | 49.15 | 82.69  | 108.43 |        | X <sup>2</sup> p-value<br>5.38378E-11 |
| Exp.           | 70.15  | 28.59  | 34.53  | 13.85 | 23.31  | 30.57  |        |                                       |
| X <sup>2</sup> | 0.19   | 1.52   | 3.78   | 4.07  | 2.83   | 0.11   | 12.51  |                                       |
| X <sup>2</sup> | 0.67   | 5.39   | 13.42  | 14.44 | 10.06  | 0.39   | 44.36  |                                       |
|                |        |        |        |       |        |        | TOTAL  | 56.87                                 |

Table XVIII is an example of a 2 x 6 Chi Square Goodness of Fit statistic test which does not reject the null hypothesis for frame ACT; the two laboratories may produce the same distribution of patterns for Frame ACT using Primer C1.

**Table XVIII** — An example of a 2 x 6 Chi Square Goodness of Fit test for independence not rejected (Appendix III, C1, same chemistry). The frame ACT yields similar relative frequencies within the patterns as can be seen by comparing the observed to the expected and Chi Square contributions between the two laboratories. The  $H_0$  is not rejected because the calculated  $X^2 = 5.58$  which is less than 21.8322 with  $p = 0.349$  which is greater than 0.00078125. Hence, the distribution of patterns may be assumed to be independent between the two laboratories. Also note that there are three expected observations (i.e., 3/12 or 25% of the observations) less than 5.0 and their contribution is minimal; therefore, it is legitimate to use the Chi Square Goodness of Fit test.

| Pattern        | A     | B     | C     | D      | E    | F     | TOTALS |                                       |
|----------------|-------|-------|-------|--------|------|-------|--------|---------------------------------------|
| ACT            | 26.00 | 63.00 | 77.00 | 163.00 | 2.00 | 8.00  | 339.00 |                                       |
| ACT            | 13.00 | 14.00 | 18.00 | 49.00  | 0.00 | 4.00  | 98.00  |                                       |
| TOTAL          | 39.00 | 77.00 | 95.00 | 212.00 | 2.00 | 12.00 | 437.00 |                                       |
| Exp.           | 30.25 | 59.73 | 73.70 | 164.46 | 1.55 | 9.31  |        | X <sup>2</sup> p-value<br>0.349082834 |
| Exp.           | 8.75  | 17.27 | 21.30 | 47.54  | 0.45 | 2.69  |        |                                       |
| X <sup>2</sup> | 0.60  | 0.18  | 0.15  | 0.01   | 0.13 | 0.18  | 1.25   |                                       |
| X <sup>2</sup> | 2.07  | 0.62  | 0.51  | 0.04   | 0.45 | 0.64  | 4.33   |                                       |
|                |       |       |       |        |      |       | TOTAL  | 5.58                                  |

For those frames with expected values less than that required in more than 25% of the cells, a Kolmogorov-Smirnov test was performed. The tables are not presented here but an example is given in Table XIX; the p-values calculated with the Kolmogorov-Smirnov test statistic were generally consistent with those of the Chi Square Goodness of Fit test. A total of 154 of the 246 frames were subjected to the Kolmogorov-Smirnov test.

**Table XIX** — Example of Kolmogorov-Smirnov Test (Appendix III, C1, same chemistry). When more than 25% of the cells had expectations less than five or when less than 25% of the cells were less than five and contributed significantly to the calculated Chi Square, the Kolmogorov-Smirnov Test, for Two Samples of size n and m, was used. Although Kolmogorov-Smirnov Test is less powerful, the results were generally consistent with the Chi Square Goodness of Fit test. CRF is the cumulative relative frequency.

| Pattern                    | A     | B     | C     | D     | E     | F     | TOTAL  | Compared to X <sup>2</sup> p-value |
|----------------------------|-------|-------|-------|-------|-------|-------|--------|------------------------------------|
| GGG                        | 1.00  | 11.00 | 0.00  | 1.00  | 0.00  | 73.00 | 86.00  | 0.816549782                        |
| GGG                        | 1.00  | 1.00  | 0.00  | 0.00  | 0.00  | 18.00 | 20.00  |                                    |
| <b>TOTAL</b>               | 2.00  | 12.00 | 0.00  | 1.00  | 0.00  | 91.00 | 106.00 | Sqrt((m+n)/mn)=<br><b>0.248</b>    |
| CRF                        | 0.012 | 0.140 | 0.140 | 0.151 | 0.151 | 1.00  | KS     |                                    |
| CRF                        | 0.050 | 0.100 | 0.100 | 0.100 | 0.100 | 1.00  | 0.266  | <b>p&gt;0.2</b>                    |
| <b> DIFF </b>              | 0.038 | 0.040 | 0.040 | 0.051 | 0.051 | 0.000 | 0.051  |                                    |
| KS = Maximum Difference  = |       |       |       |       |       |       |        |                                    |

In summary, the null hypothesis of independence of the distribution of patterns between Laboratory A and Laboratory B is rejected for 77 of 246 frames. It is not rejected for 169 of 246 frames. Ten frames have no observations. The fact that ten of the frames have no observations (the three consecutive bases in that frame does not exist for this region) is extremely important and can be programmed into the expert system for flagging if that frame were to appear. This could suggest that there is an erroneous result or it could suggest a very rare event.

These results demonstrate that the data obtained from the two laboratories are comparable but not the same. Each of the four primers, A1, B1, C1, and D1, were evaluated to characterize the different patterns for each frame. With these results, an expert system could be initially programmed with a single laboratory’s parameters and then optimized for another laboratory with its own sequence data.

### 3.8.4 — Statistical significance of pattern distribution between two laboratories

Appendix IV uses the same tools to statistically evaluate each of the 64 frames using a 2 x 6 Chi Square analysis spreadsheet, comparing the results from data from different primers produced in the same laboratory. The results in Appendix IV demonstrate that the data obtained from the different primers are not comparable. Comparisons were made between Primers A1 and C1; Primers B1 and D1; and Primers A1 and B1. Data for each frame for the different primers cannot be pooled into a single database.

---

Appendix V also uses the same tools to statistically evaluate each of the 64 frames using a 2 x 6 Chi Square analysis spreadsheet, comparing the results from data from different dye chemistries for sequences produced in the same laboratory. The results in Appendix V demonstrate that the data obtained from the different dye chemistries are not comparable. Comparisons were made between BigDye v1.1 dye chemistry and dRhodamine.

### 3.9 — Patterns can be predicted

All results suggest that, if the chemical parameters and primers used are kept constant, patterns can be predicted for a given laboratory in future sample runs.

This means that one can build a pattern reference table for each frame given the primer and chemistry used in the sequencing. We call this the Sequence Biometrics standard table for this chemistry/primer combination.

An example of such table is depicted in Table XX.

**Table XX** (next page) – Example of the frequency table of patterns for each frame in Laboratory B, using BigDye 1.1, chemistry and primer C1. Gray cells are less than 10% frequent (rare) and green cells are over 50% frequent or more frequent for that frame

Characterizing patterns in DNA sequence trace data through informatics tools

|     | A      | B      | C      | D      | E      | F      | TOTAL   | p-Value       |
|-----|--------|--------|--------|--------|--------|--------|---------|---------------|
| AAA | 20.90% | 2.54%  | 46.87% | 0.15%  | 16.72% | 12.84% | 670.00  | 7.3393E-121   |
| AAC | 35.72% | 10.01% | 14.22% | 0.46%  | 33.11% | 6.48%  | 879.00  | 1.8946E-117   |
| ACA | 9.19%  | 42.10% | 1.25%  | 29.65% | 8.70%  | 9.12%  | 1437.00 | 6.0921E-227   |
| ACC | 33.94% | 12.48% | 20.20% | 4.91%  | 13.74% | 14.73% | 713.00  | 2.5634E-42    |
| ACT | 7.67%  | 18.58% | 22.71% | 48.08% | 0.59%  | 2.36%  | 339.00  | 6.09762E-67   |
| ACG | 0.29%  | 5.57%  | 0.00%  | 65.98% | 0.00%  | 28.15% | 341.00  | 5.449E-153    |
| AAT | 2.06%  | 12.79% | 1.25%  | 11.00% | 30.32% | 42.58% | 1118.00 | 2.2824E-194   |
| ATA | 34.24% | 5.71%  | 37.10% | 1.12%  | 9.06%  | 12.78% | 806.00  | 4.0275E-119   |
| ATT | 14.46% | 0.74%  | 24.79% | 0.58%  | 51.07% | 8.35%  | 1210.00 | 4.0832E-286   |
| ATC | 6.17%  | 16.30% | 0.69%  | 28.47% | 2.06%  | 46.31% | 583.00  | 1.6738E-118   |
| ATG | 66.51% | 1.17%  | 30.21% | 0.00%  | 1.87%  | 0.23%  | 427.00  | 2.8625E-201   |
| AAG | 11.48% | 10.93% | 3.83%  | 49.18% | 1.64%  | 22.95% | 183.00  | 7.71945E-35   |
| AGA | 21.07% | 0.53%  | 74.13% | 2.67%  | 0.27%  | 1.33%  | 375.00  | 4.8407E-206   |
| AGG | 10.53% | 0.44%  | 88.16% | 0.00%  | 0.88%  | 0.00%  | 228.00  | 1.3894E-181   |
| AGC | 1.17%  | 0.00%  | 59.43% | 0.50%  | 16.53% | 22.37% | 599.00  | 6.3073E-203   |
| AGT | 11.06% | 2.13%  | 35.74% | 0.43%  | 50.21% | 0.43%  | 235.00  | 9.97307E-67   |
| CAA | 0.28%  | 24.51% | 0.70%  | 6.41%  | 36.49% | 31.62% | 718.00  | 1.9824E-119   |
| CAC | 11.20% | 18.55% | 30.71% | 0.44%  | 30.53% | 8.57%  | 1143.00 | 3.4639E-109   |
| CCA | 0.90%  | 8.45%  | 3.92%  | 58.52% | 7.54%  | 20.66% | 663.00  | 4.4458E-198   |
| CCC | 0.90%  | 3.85%  | 0.23%  | 70.36% | 1.36%  | 23.30% | 442.00  | 3.4124E-218   |
| CCT | 7.27%  | 17.04% | 2.11%  | 25.23% | 17.70% | 30.65% | 757.00  | 6.49566E-54   |
| CCG | 8.23%  | 34.49% | 19.62% | 0.63%  | 3.80%  | 33.23% | 316.00  | 6.86065E-43   |
| CAT | 14.71% | 24.55% | 2.69%  | 8.06%  | 36.70% | 13.30% | 782.00  | 1.0019E-73    |
| CTA | 8.39%  | 8.39%  | 46.40% | 6.16%  | 20.03% | 10.62% | 584.00  | 4.00094E-87   |
| CTT | 40.20% | 8.26%  | 19.75% | 0.42%  | 22.83% | 8.54%  | 714.00  | 1.57331E-90   |
| CTC | 3.18%  | 9.89%  | 1.77%  | 41.34% | 2.83%  | 40.99% | 283.00  | 1.82576E-65   |
| CTG | 59.38% | 0.45%  | 23.81% | 0.30%  | 12.20% | 3.87%  | 672.00  | 1.6213E-223   |
| CAG | 2.02%  | 1.87%  | 0.29%  | 19.02% | 3.17%  | 73.63% | 694.00  | 2.8300E-370   |
| CGA | 0.00%  | 0.00%  | 66.67% | 0.00%  | 33.33% | 0.00%  | 327.00  | 1.1667E-162   |
| CGG | 25.88% | 6.47%  | 55.88% | 0.00%  | 10.59% | 1.18%  | 170.00  | 2.81304E-48   |
| CGC | 1.85%  | 2.16%  | 34.31% | 28.13% | 16.38% | 17.16% | 647.00  | 4.44939E-71   |
| CGT | 2.65%  | 1.99%  | 78.81% | 0.66%  | 15.23% | 0.66%  | 151.00  | 1.54625E-91   |
| TAA | 7.67%  | 28.45% | 4.51%  | 47.02% | 0.34%  | 12.01% | 1174.00 | 6.0223E-238   |
| TAC | 32.22% | 2.30%  | 54.60% | 1.26%  | 7.95%  | 1.67%  | 478.00  | 4.2511E-148   |
| TCA | 0.47%  | 0.00%  | 68.15% | 0.00%  | 31.38% | 0.00%  | 427.00  | 2.9824E-217   |
| TCC | 3.04%  | 2.02%  | 18.02% | 6.28%  | 40.89% | 29.76% | 494.00  | 4.67588E-79   |
| TCT | 5.39%  | 4.11%  | 32.06% | 13.90% | 37.73% | 6.81%  | 705.00  | 1.31091E-95   |
| TCG | 0.38%  | 0.00%  | 4.91%  | 4.53%  | 40.75% | 49.43% | 265.00  | 4.01349E-83   |
| TAT | 13.63% | 45.38% | 16.59% | 18.13% | 5.45%  | 0.83%  | 844.00  | 1.9361E-130   |
| TTA | 37.98% | 16.19% | 37.17% | 3.26%  | 1.43%  | 3.97%  | 982.00  | 3.4131E-182   |
| TTT | 4.35%  | 52.95% | 0.98%  | 36.52% | 0.42%  | 4.78%  | 712.00  | 6.6018E-230   |
| TTC | 5.23%  | 17.08% | 0.62%  | 12.46% | 0.46%  | 64.15% | 650.00  | 2.4707E-244   |
| TTG | 49.72% | 2.78%  | 41.74% | 1.67%  | 1.11%  | 2.97%  | 539.00  | 2.6052E-177   |
| TAG | 0.00%  | 0.46%  | 0.92%  | 58.99% | 1.38%  | 38.25% | 217.00  | 4.75852E-90   |
| TGA | 3.86%  | 78.97% | 1.29%  | 11.59% | 1.72%  | 2.58%  | 233.00  | 1.1671E-140   |
| TGG | 5.59%  | 52.35% | 1.47%  | 27.06% | 2.06%  | 11.47% | 340.00  | 6.65475E-85   |
| TGC | 9.30%  | 36.77% | 0.73%  | 51.89% | 0.15%  | 1.16%  | 688.00  | 6.9029E-218   |
| TGT | 0.00%  | 20.08% | 0.13%  | 79.54% | 0.00%  | 0.26%  | 782.00  | < 1%0048E-499 |
| GAA | 0.67%  | 54.21% | 0.00%  | 8.75%  | 0.00%  | 36.36% | 297.00  | 1.1894E-100   |
| GAC | 9.80%  | 57.42% | 0.56%  | 31.93% | 0.00%  | 0.28%  | 357.00  | 6.5042E-125   |
| GCA | 19.68% | 27.68% | 18.70% | 8.12%  | 0.12%  | 25.71% | 813.00  | 2.90678E-57   |
| GCC | 12.98% | 20.70% | 18.07% | 0.70%  | 43.68% | 3.86%  | 570.00  | 4.5238E-85    |
| GCT | 1.86%  | 25.99% | 40.60% | 2.09%  | 1.39%  | 28.07% | 431.00  | 3.79508E-79   |
| GCG | 2.70%  | 1.89%  | 7.28%  | 29.38% | 1.62%  | 57.14% | 371.00  | 2.216E-119    |
| GAT | 0.74%  | 56.25% | 0.00%  | 2.94%  | 0.00%  | 40.07% | 272.00  | 1.5307E-107   |
| GTA | 4.86%  | 29.79% | 9.73%  | 1.22%  | 9.73%  | 44.68% | 329.00  | 5.65438E-59   |
| GTT | 0.00%  | 48.77% | 0.00%  | 0.00%  | 4.51%  | 46.72% | 244.00  | 5.02487E-90   |
| GTC | 6.82%  | 1.31%  | 0.52%  | 0.00%  | 23.88% | 67.45% | 381.00  | 8.3542E-171   |
| GTG | 28.87% | 0.26%  | 20.36% | 0.00%  | 50.26% | 0.26%  | 388.00  | 8.8089E-104   |
| GAG | 22.58% | 8.50%  | 0.29%  | 12.61% | 1.47%  | 54.55% | 341.00  | 1.56937E-88   |
| GGA | 35.94% | 29.28% | 3.19%  | 7.83%  | 4.93%  | 18.84% | 345.00  | 8.47539E-40   |
| GGG | 1.16%  | 12.79% | 0.00%  | 1.16%  | 0.00%  | 84.88% | 86.00   | 1.62366E-61   |
| GGC | 2.04%  | 42.45% | 0.00%  | 0.82%  | 8.98%  | 45.71% | 245.00  | 2.96643E-71   |
| GGT | 0.59%  | 8.82%  | 0.00%  | 72.94% | 1.76%  | 15.88% | 170.00  | 1.06385E-85   |



Hence, it is possible to predict the behavior of patterns for any given frame if all the parameters between samples are kept constant within the same laboratory. This prediction can also be made using samples that were run in different instruments or at different timeframes within the laboratory. It is proven that changing the primer or chemistry will change, within the laboratory, the distribution of patterns. All in all, frame pattern tables can be a tool for calibrating and predicting frame pattern results for any sample within the laboratory, given its run parameters on that laboratory. Besides, building these tables can allow software tools to predict future patterns in sequence traces and to understand comprehensively changes in these files.







---

## 4.1 — Presumptions

First and foremost, we must start by defending all the assumptions that led to the origin of the study, as well as some factors that were taken for granted first hand.

As described in the introduction there were a few hints and results that indicated that peak height patterns meant something and might not be random for some frames. This became clear with our investigation and it could be proved that those assumptions were true, yet underestimated, of the real phenomenon.

This study assumes that several technical parameters are held constant throughout this investigation, which is debatable, but nevertheless we would not be able to study this phenomenon if we did not assume that chemistry parameters are kept mostly constant for the same kind of samples and primers, that those primers are the same or very similar either from laboratory to laboratory or in forward and reverse sequencing and that methodology within the same laboratory is consistent throughout time. Furthermore, it is assumed that any software will interpret .AB1 files in the same way. This is not necessarily true for the first basecalls of any sequence, so in our comparisons we always took out the first 5-10 bases. We call that an artifact but in fact it is more of a variation of interpretation of sequence trace data, at the beginning of the trace, done by different basecallers.

Hence, three major presumptions were made and kept throughout this study:

1 – The same chemistry means that there are minor variations in the chemical or technical parameters used to obtain the trace sequence from the sample. One cannot assume that methodology of one laboratory is exactly the same as methodology of another laboratory. In this study, standard protocols were used in each laboratory. When comparing results from one laboratory to another, it cannot be presumed that all procedural steps are equivalent. This study shows that differences between laboratories' methodologies, machines, reagents and samples each produce a unique distribution of patterns within frames for the laboratory.

2 – Different samples processed in the same laboratory with the same parameters, such as chemistry or primers, are only as variable as the sample itself. It is assumed that neither methodologies nor chemistry or primers have changed in the laboratory. That is, it is assumed that

any variation within two samples processed in the same laboratory occurs because of differences between the samples and not because of the moment when samples were processed or the batch of chemicals used in the reaction. Within the laboratory it is assumed that everything but the sample is kept constant when comparing two samples processed with the same chemistry and sequenced with the same primer.

3 – Different software programs will output the same relative heights for the same .AB1 file and will interpret sequence trace data in a similar manner. Hence different software programs will produce similar results. Sometimes, especially at the beginning and ends of sequences traces with low quality data, different algorithms within each program will interpret data differently.

## 4.2 — Methods

As stated in the Materials and Methods section, peak height patterns are dependant on the resolution of the sequencer. Two or three peaks within a frame can have the same height value. When two or three peaks have the same height, pattern grouping was defined. For this study, pattern grouping was not considered in the overall characterization of patterns for each frame because they are rare. Hence, though their existence was not ignored, the overall weight of groups within each frame was so low that they were not used to characterize the overall pattern distribution on any of the frames, but they were used when comparing two or more similar sequences side-by-side.

Flip-flopping, another common phenomenon, was also addressed by characterizing minor and major differences. Yet, we never set a threshold on flipping and believed that flip-flopping would not occur with significant peak height differences between the two flipping peaks. That is, of course, a simplification. There is the possibility of a minor difference becoming a major one if a set threshold is determined and the change is beyond the limits of that threshold and also, a major change could become minor if below the defined threshold. Now, the reason why thresholds were not defined has to be cleared. What would be a perfect threshold? An exact value or a percentage? Why choose 5% of the total peak height as the threshold and not 10%? Such questions led us to focus on the phenomenon of flip-flopping without looking at all the heights of the peaks involved and calculating debatable thresholds.

---

These decisions slightly affect the results but, in our point of view, such effect is not weighted enough to affect the overall results and conclusions. Hence, whether groups and thresholds are considered or not, we are convinced that the overall results and conclusions will not change dramatically.

Secondly, there is also the reason why we chose frames of three bases and not more or less bases. Simplicity was the key, as three bases allow for 64 frame combinations. Yet, one base height tells nothing and two bases in a frame cannot say more than which height is smaller and which is bigger. To get the most information with the least possible frame combinations, frames of three bases were chosen because they allow for the extraction of more information out of all combinations of peak heights for the smallest number of possible peaks. If patterns occur within these three base frames, then they might also occur for frames of larger number of bases, but calculating and most importantly understanding that data would be a task requiring super-computing capabilities. If frames of three bases have 64 combinations and each frame can exhibit six different patterns (excluding groups), which is the equivalent to 384 possibilities. A frame of four bases has 256 combinations and can exhibit 12 different patterns (again, excluding groups), which is the equivalent to 3072 possibilities. It is easy to understand that three bases within a frame provide some information for our study and that future studies may want to evaluate larger frames.

Another interesting observation was non-independent data for calculating patterns. Non-independent frames occur when the heights of the peaks in that frame are given or altered by the height of any of the peaks of the proceeding frame due to homopolymeric stretches. Hence, the latter frames are not statistically independent of the previous ones because their peak heights are affected by the heights of other peaks in earlier frames. When this happens, the outcome pattern of that frame is not an independent event and therefore is not considered to have a statistical significance; therefore, it should be removed from our analysis since it adds randomness. This is a common occurrence in the Sanger sequencing method when dealing with homopolymeric stretches. Here, the first frame of three repeats is considered but the next repetitive frames are not taken into account because their heights are dependant on the heights of the first homopolymeric frame.

In other sequencing procedures, such as in pyrosequencing, this phenomenon is much more serious because base signals are given as an average. In this kind of sequencing a repeat of equal bases will have a certain signal that is given by the mathematical average of the total repeat signal. Hence, the signal of the base will be dependant on its intensity and of the neighboring bases. The

extension of this behavior is so important that, even though we had access to data other than Sanger sequencing data, we were not able to use these data because of severe independence issues. Yet, the extension of the independence problem made it too difficult to treat these data and we focused our study on the Sanger sequencing data.

As was said before, we assume that any basecaller will make its choices and call bases in a similar way and that all sequences will look similar from application to application. That is why editing files with Finch TV or 4Peaks produces sequences similar to the ones exported through Mutation Surveyor<sup>®</sup>. This is not such a straightforward process and, in fact, some minor differences do occur. That is why the first bases are never considered, as explained before. When trimming occurs in the sequence, different basecallers will analyze the first data points differently and try their best to provide an accurate basecall. Trimming may result in different basecalls being made, at the beginning of sequences, by different basecallers. Usually after a few basecalls everything goes back to normal and different basecallers output the exact same bases for the sequence trace. This phenomenon also occurs when the basecaller cannot provide an accurate call on a low quality peak. Different basecallers use different algorithms that try to cope with ambiguous data and they do it differently. It is now obvious that editing was needed, for this first analytical study of peak height patterns, since only the best stretches of sequence data were used to produce these results, therefore minimizing the possibilities of base call ambiguity.

There is also much to be said about the Python scripts used for finding and counting patterns in the data files. “Sanger”, “Cleaner” and “Reporter” were the three main scripts used but they show several minor limitations. The most important script is the “Sanger” script because it not only finds patterns within frames but also does it using its own basecalling, by comparing the signal intensity of each of the four dyes at the basecall point. Sometimes this new basecall differs from the basecall made by Mutation Surveyor for reasons that we have explained before, especially in low quality basecalls. Though such issues are rare and insignificant because of the quality of the data used, they do happen from time to time. Also, “Sanger” ignores all of the quality values from the basecall because it was not designed to provide this feature. We think that the quality assessment from 4Peaks or FinchTV is enough and further editing downstream in our workflow, during the script run is not usually necessary.

“Cleaner” is a simple script designed to replace frame data after the first homopolymeric frame in a stretch of bases and stops cleaning when the first non-homopolymeric frame appears



downstream. Its major limitation is that it is difficult to add new parameters to the cleaning procedure. In this way, it is very specific for the Sanger sequencing procedure.

“Reporter” is also simple in nature, but its implementation makes it the slowest script that we used. Although there were some options to count the patterns for each frame, we decided to use a simple but blunt approach to counting. Our “Reporter” script uses a counter for each and every expected frame pattern occurrence, reads the data file from top to bottom for any given counter, stores the count and repeats this procedure to the next counter. With over seven hundred counters, “Reporter” is memory intensive and slow yet very simple. This allows for faster changes and easier debugging than any other solution, which is a curious fact since the program has over 2000 lines of source code. These Python scripts were kept separate for simplicity. Although all the calculations could be performed using a single script, I chose to keep the scripts independent to perform their tasks individually for easier debugging, better understanding and faster modification, if necessary.

Excel<sup>®</sup> spreadsheets also follow the same simplicity philosophy. They are complex in size yet most formulae in the spreadsheet are the combination of simple Boolean functions for each cell. This is also important for debugging the spreadsheet or modifying its contents. Excel<sup>®</sup> statistical calculations are also simple and common. Though other statistical programs could be used, the simplicity of this application, the ease of programming the spreadsheets, testing, making collaborative changes and its capability in handling these calculations so well, made us choose it rather than any other program.

All in all, these spreadsheets were a simple way to calculate the distribution of the pattern findings. Yet, as it later became clear, a simple Chi Square Goodness of Fit test was sufficient for most of the calculations. With few observations, the Kolmogorov-Smirnov test was performed. The results from the Kolmogorov-Smirnov tests generally agreed with the Chi Square results. That is, for the general conclusions of our work, a simple Chi Square Goodness of Fit test performed in an Excel<sup>®</sup> spreadsheet is sufficient; though, we have shown that more advanced tests can be performed on the data in order to improve the reliability of the results.

Again, Excel<sup>®</sup> was used for simplicity, availability and development speed but it is not, in our opinion, the best tool to execute the statistical tests. We hope, in the future, to develop new ways of performing such calculations. Perhaps we shall use Python because, in fact, the first draft of a pattern finder script was built on Excel<sup>®</sup> and then later transported into this programming language. This was possible because the simple calculations in the spreadsheets can be translated

into scripting code without much modification. Hence, the statistics spreadsheets could also be, in the future, translated into a language such as Python or any other streamlined programming language.

### 4.3 — Results

Up to this point we discussed our choices and decisions to analyze and characterize patterns in trace sequence frames. We shall start by answering the first question: Are patterns kept constant if a sample is run twice on the sequencer? If a sample run twice on the same instrument were not to exhibit the same distribution of patterns in each run, the pattern formation would be considered random. Such is not true and we demonstrated this by doing the simplest test, a side-by-side comparison of two electropherograms. As it becomes clear after repeating the process, sequences are, in fact, similar when the same sample is run twice on the same instrument. This was expected, yet our results show beyond any doubt that this is the truth. Patterns are kept constant within the frames from one run to the other. Yet, though sometimes differences occur, they are mostly minor and dependant on the resolution of the equipment.

Two other questions sprouted naturally once we understood that patterns are not random. Do two very similar samples exhibit the same patterns for each frame? Yes. Do the same samples run on two different instruments produce similar patterns? Absolutely. These outcomes are of extreme importance. Not only are patterns kept constant if the same sample is run twice independent of the instrument used, but also similar samples, run with the same chemistry, sequenced with the same primers and processed in the same laboratory will show similar patterns for each frame.

The positive response for these three questions led us to the fourth, and perhaps the most important question of the study: Can similar sequences be pooled together and will their frames show any tendency for a type of pattern? This question was answered in section 3.4 where it was made clear that not only sequences can be pooled together, but the sum of those sequence frame patterns will exhibit, for the majority of frames, a non-random behavior. That is, when we pool sequences together for the same chemistry, primer and laboratory, the resulting pool of frames will show definable and predictable patterns for any type of pattern, independent of the instrument used in the laboratory.

So, not only patterns are kept constant in the frames from sequence run to sequence run, but also from sample to sample if the conditions are the same. This happens independent from the instrument. Also, these sequence traces are so similar that when pooled together they will form predictable patterns for each frame. Such a result is fundamental for any further study of patterns and this is the first time patterns within frames are not only thoughtfully studied but also can be predicted for each and any frame using simple computer programs.

This new method of qualitative assessment tools for sequences data has never, to date, been explored. By knowing which patterns a certain frame may exhibit and most importantly which patterns it will never exhibit, programs can be written to signal the analyst to rare sequence results. It is possible to calibrate and to have a standard table of the distribution of frame patterns for DNA sequence data using the Sanger method with fluorescence dye chemistry. Hence we have shown that peak height patterns are an intrinsic characteristic of the DNA sequencing procedure; we have developed methods for interpreting and predicting them.

The next analysis after this point was to understand the distribution of peak height patterns within frames of sequence trace data. If sequences run on two different instruments are similar and different samples with all parameters constant also show comparable patterns, do two laboratories, sequencing the same DNA regions, using equivalent techniques, procedures and instruments produce similar distribution of patterns? Our results show that there are some similarities in the distribution of patterns between laboratories. Sequence frame pattern distributions are globally similar from laboratory to laboratory if the chemistry and primer combination is the same. It is clear that peak height patterns within frames in sequence trace data are best conserved in the same laboratory, using the same standardized methods. Using the results of pattern distribution from one laboratory to the other is too much of a risk because of all the subtle, yet important, differences in methodology and processing that might affect the overall pattern distribution. Yet, global chemistry/primer combination results are a good start for the optimization of software definitions of sequence patterns in other laboratories. All in all, though the peak height patterns for a chemistry/primer combination of Laboratory A could, theoretically, be merged with the patterns of Laboratory B, in practice, we think it is best for each laboratory to be treated independently, as if it was producing different patterns.

So, if patterns are not highly conserved from one laboratory to the other for similar samples run with similar procedures, what is affecting the pattern distribution? There are two main reasons

that can change the way patterns are formed. One is the DNA sequence itself and the other is the dye chemistry used. Therefore, we initiated a study to determine if the same samples run in the same laboratory and instrument but sequenced with different primers, hence sequencing different DNA areas, would show the same pattern distribution. If the pattern distribution were similar, then the DNA sequence itself would not affect the distribution of patterns. Patterns within frames cannot be compared when using different primers. If the sequence is different, then the frames of that sequence will show different pattern distribution.

Later, we also tried to understand if the chemistry would affect pattern distribution. Our results show that different chemistries produce a different distribution patterns for most frame. The distributions of peak height patterns for a frame within a sequence are an intrinsic characteristic of that sequence and that procedure. But these patterns are repeatable and can be predicted if the parameters are known and kept constant.

Sequence Biometrics are the qualitative values defined for sequence data. Different DNA sequence traces have a defined distribution of patterns for each frame and similar sequence traces display similar distributions. Yet, when the chemistry varies so does the distribution of patterns and if the sequenced DNA region changes, so does the distribution. The distribution of patterns within a sequence frame is not random; the distribution is characteristic of the DNA region and the sequencing procedure. Sequence Biometrics allows the analyst to build standard pattern distribution tables to describe its frames for a specific DNA region and chemistry and use this calibration to better understand changes for similar sequences. If the sequencing procedure is standardized, such as it is in a forensic laboratory, the analyst can build a calibration table for each primer and chemistry using very few samples. The analyst can compare the patterns for each frame processed and then be sure that any other sample, processed similarly within the laboratory, will exhibit similar peak height patterns for each frame. If different, then the analyst can study the data closely.

We provide in section 3.9 a first example of a standard pattern frequency table for each frame of a specific DNA area sequenced using specific dye chemistry. Most, if not all of the frames, show predominant patterns within each frame. Moreover, the distribution is different from frame to frame and, as said before, not random. Hence, by building a prediction table for the sequence, that is the Sequence Biometrics, one can more easily find aberrances in each of the frames throughout the sequences.

The calibration must be done within the laboratory, for a specific chemistry/primer combination. Although these biometric values can be transported to other laboratories using similar methodologies, it is our opinion that each laboratory should build their own standard sequence biometric tables using the tools that I have designed.

#### **4.4 — Sequence Biometrics**

During our study, it has become evident that each sequence trace is highly repeatable and can be compared with sibling sequences that share characteristics such as the DNA area sequenced and sequencing chemistry. Hence, we have unraveled a new characteristic of DNA sequence traces: Sequence Biometrics. Sequence Biometrics is the parameter that allows a sequence trace to be compared with other similar sequence traces. We can separate populations of sequence traces according to their DNA area sequenced and dye chemistry used but we cannot compare different populations. It is clear that sequences within the same population are comparable and that such population exhibits definable and conserved patterns within each frame, which are dependant on the context of the sequence.

In our opinion, there are at least two factors that, when combined, form these different populations. First, the sequence template, or DNA area sequenced, will affect the amount of dye that is incorporated at a given time with the cycle sequencing procedure used. Second, the chemistry used will affect the different dye incorporation rates since not all dyes take the same time to be incorporated into the PCR product. These two factors, when combined, can output a much bigger change in the overall biometrics as is evident in our results.

If these were the two sole factors that had anything to do with Sequence Biometrics, DNA area and dye incorporation rate, then we could compare populations of two different laboratories. But this is not true. The reaction environment is a third factor that affects Sequence Biometrics. Different laboratories use slightly different procedures that can change the dye incorporation kinetics in each laboratory. Hence, Sequence Biometrics is dependant on a multitude of variables, some more significant than others but, when combined, provide an identity for each sequence in each laboratory.

If analytical parameters and DNA sequences are kept constant, so will the Sequence Biometrics. So, if Laboratory A and Laboratory B were to use the exact same extraction procedure, PCR cycling and chemistry, standardizing the procedures and diminishing variability, the biometrics of each DNA area sequenced in both laboratories could be the same and highly comparable, because we have shown that Sequence Biometrics is not dependant on the sequencing instrument. All in all, by knowing the Sequence Biometrics for a DNA area processed with a known chemistry, we can predict the patterns in each frame for a new sequence that has been sequenced using the same parameters; therefore, new Sequence Biometrics prediction is also possible if properly controlled.

#### **4.5 — Future development**

From what we have been discussing so far, it is clear that the use of patterns in future sequencing applications is not only possible but will also add value to the entire process. It has become clear that sequences exhibit discrete and predictable frame patterns for any type of dye chemistry and primer combination but that those patterns are characteristic of such parameters. Hence, Sequence Biometrics defines what patterns will be more common for a sequence when sequenced with a determined chemistry in a determined laboratory that applies standardized sequencing procedures. The existence and properties of such Sequence Biometrics allows a user to calibrate a table of pattern frames for each situation and predict, in future runs, the pattern output for the sequences it is using. This not only provides a better quality assessment for the analyst but also a faster and more reliable way of interpreting sequence data, and most importantly, finding aberrations.

At the end of our methodology, we proposed a new type of system for looking at sequences: an expert system. The expert system that we propose must be able, having been fed with a standard pattern table for each different laboratory situation, to validate and interpret future sequences that use the same method and chemistry. Our investigation defines the theoretical basis for such a system, provides some of the basic procedures for automated pattern distribution characterization and answers some of the fundamental questions for understanding the mechanisms inherent to patterns.

In the future, we plan on conducting a thorough study on mutations and mixtures of DNA and the effect on the patterns observed. It is clear, especially when looking at differences from sequence to sequence, that such variations will alter the overall distribution of patterns in the frames that are affected by it. Such differences could be flagged by the expert system. All in all, we propose a more automated and reliable way to look at sequences using Sequence Biometric information (such as distribution of patterns in a frame) to provide further quality assessments in future integrated expert systems.

#### **4.6 — Final thoughts**

Thus, throughout this study we have been able to prove that, in sequence trace data, peak height patterns within frames are not random and can be predicted. Moreover, we have shown that each DNA sequence area, dye chemistry and laboratory procedures they produce a repeatable batch of pattern distributions for a frame. Hence, Sequence Biometrics can be applied for each laboratory by knowing the type of chemistry used and the DNA area sequenced. These biometric parameters are specific for the DNA area sequenced but are only kept constant if the sequencing parameters are the same, independent of the instrument used to read the samples. In conclusion, we have set the foundation for a new type of computer tool to analyze sequence trace data: the expert system with pattern finding and Sequence Biometrics characterization abilities.









From our study, several conclusions can be made from our results.

- Tools such as Mutation Surveyor® that allow the export of sequence trace data provide qualitative data for further analysis
- By examining several software tools for our study, we were able to determine which features should be present on a future expert system
- Peak height patterns in DNA sequence trace data frames are not random and can be characterized and predicted
- Peak height patterns are conserved if the chemistry and primer combination used to sequence DNA are kept constant from sample to sample
- Different instruments do not affect peak height pattern distribution
- Pattern distributions are best conserved within the same laboratory by keeping all the parameters the same
- Different laboratories produce similar, but not equal, pattern distributions for each chemistry/primer combination because of small variations in protocols
- The distribution of peak height patterns defines a new characteristic of sequence trace data: Sequence Biometrics
- Sequence Biometrics standard tables can be built for each chemistry/primer combination group within the laboratory and define the characteristic pattern distribution for each of those groups
- Sequence Biometrics is a new quality parameter for sequence analysis and can be built into new expert systems
- New expert systems should be able to benefit from Sequence Biometrics to identify alterations in DNA sequence trace files



---

## References



- 
- Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, & Young IG (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465
- Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, & Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23: 147
- Bogenhagen D & Clayton DA (1974) The number of mitochondrial deoxyribonucleic acid genomes in mouse L and HeLa cells: Quantitative isolation of mitochondrial deoxyribonucleic acid. *J Biol Chem* 249: 7991-7995
- Brown WM, George M, Jr, & Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* 76: 1967-1971
- Budowle B, Smith J, Moretti T, & DiZinno J (2000) *DNA Typing Protocols: Molecular Biology and Forensic Analysis*. Eaton Publishing: Natick, MA
- Cann RL, Stoneking M, & Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325: 31-36
- Copeland WC (2002) Mitochondrial DNA. Methods and protocols. *Methods Mol Biol* 197: v-vi
- Englund PT (1972) The 3'-terminal nucleotide sequences of T7 DNA. *J Mol Biol* 66: 209-224
- Englund PT (1971) Analysis of nucleotide sequences at 3' termini of duplex deoxyribonucleic acid with the use of the T4 deoxyribonucleic acid polymerase. *J Biol Chem* 246: 3269-3276
- Ewing B, Hillier L, Wendl MC, & Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 8: 175-185
- Giles RE, Blanc H, Cann HM, & Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77: 6715-6719
- Gray MW (1992) The endosymbiont hypothesis revisited. *Int Rev Cytol* 141: 233-357
- Grivell LA (1997) Mitochondria. In *McGraw-Hill Encyclopedia of Science & Technology*, Anonymous McGraw-Hill: New York
- Ingman M, Kaessmann H, Paabo S, & Gyllensten U (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* 408: 708-713
- Robin ED & Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol* 136: 507-513
- Rosenthal A & Charnock-Jones DS (1992) New protocols for DNA sequencing with dye terminators. *DNA Seq* 3: 61-64
- Ruiz-Pesini E, Lott MT, Procaccio V, Poole JC, Brandon MC, Mishmar D, Yi C, Kreuziger J, Baldi P, & Wallace DC (2007) An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res* 35: D823-8
- Sanger F & Coulson AR (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J Mol Biol* 94: 441-448
- Sanger F, Nicklen S, & Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74: 5463-5467
- Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, & Hood LE (1986) Fluorescence detection in automated DNA sequence analysis. *Nature* 321: 674-679
- Tracy TE & Mulcahy LS (1991) A simple method for direct automated sequencing of PCR fragments. *BioTechniques* 11: 68-75
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M *et al* (2001) The sequence of the human genome. *Science* 291: 1304-1351
- Zakeri H, Amparo G, Chen SM, Spurgeon S, & Kwok PY (1998) Peak height pattern in dichloro-rhodamine and energy transfer dye terminator sequencing. *BioTechniques* 25: 406-10, 412-4
-









## Appendix I — Comparisons of Frames with Two Sequence Traces

The tables in this appendix are automatically generated from the bioinformatics tools designed for this thesis. The first row is the frame obtained from the sequence. The second row is the pattern from each frame for the first injection or sample. The third row is the pattern from each frame for the second injection or sample. The Consensus row displays the consensus pattern for the sequences compared.

A dash (-) in the first row signifies those frames that have been deleted due to sequence variation in the base caller program between samples in the second and third rows. Differences in the same sample are usually due to base sequence differences or dye background or other sequencing anomalies that occur with dye chemistry and capillary electrophoresis. A triple dash (---) in the first row signifies data that have been deleted due to frames exhibiting homopolymeric stretches of a particular base and cannot be considered independent. An asterisk (\*) in the Consensus row signifies a different sequence between the samples/injections in the frame or a Minor pattern difference when comparing Rows 2 and 3. A number sign (#) in the Consensus row signifies a Major pattern difference when comparing the frames, Row 2, and Row 3.

With the sequencing protocol held constant (i.e., BigDye™ sequencing chemistry v1.1, BetterBuffer, 3130xl Genetic Analyzer, POP-6), the following comparisons have been made between two injections. Each legend describes the comparisons performed with sequence metrics. Overwhelmingly, consensus of patterns is achieved.

**Table A.** The same sample sequenced twice with the A1 primer electrophoresed on the same instrument in the same run. The frames displayed in this example are from position 16,019 (p16,019) to p16,398.

|                  |    |      |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------------------|----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <b>Frame</b>     | 1  | -    | CCT | CTC | TCT | CTG | TGT | GTT | TTC | TCT | CTT | TTT | -   | -   | -   |     |
| <b>Inj.1</b>     | 1  | A    | A   | D   | C   | A   | D   | E   | B   | C   | A   | A   | D   | E   | A   |     |
| <b>Inj.2</b>     | 1  | B    | C   | D   | C   | A   | D   | E   | B   | C   | A   | A   | D   | C   | A   |     |
| <b>Consensus</b> | 1  | *    | *   | D   | C   | A   | D   | E   | B   | C   | A   | A   | *   | *   | *   |     |
| <b>Frame</b>     | 15 | TGG  | GGG | --- | GGA | GAA | AAG | AGC | GCA | CAG | AGA | GAT | ATT | TTT | TTG | TGG |
| <b>Inj.1</b>     | 15 | BETA | F   | -   | A   | A   | D   | E   | B   | F   | C   | B   | F   | F   | C   | B   |
| <b>Inj.2</b>     | 15 | D    | F   | -   | A   | A   | D   | E   | B   | F   | C   | B   | F   | F   | C   | B   |
| <b>Consensus</b> | 15 | *    | F   | -   | A   | A   | D   | E   | B   | F   | C   | B   | F   | F   | C   | B   |
| <b>Frame</b>     | 30 | GGG  | GGT | GTA | TAC | ACC | CCA | CAC | ACC | CCC | CCA | CAA | AAG | AGT | GTA | TAT |
| <b>Inj.1</b>     | 30 | F    | F   | C   | A   | D   | F   | F   | C   | B   | F   | F   | E   | A   | B   | C   |
| <b>Inj.2</b>     | 30 | F    | F   | C   | A   | D   | F   | F   | C   | B   | F   | F   | E   | A   | B   | C   |
| <b>Consensus</b> | 30 | F    | F   | C   | A   | D   | F   | F   | C   | B   | F   | F   | E   | A   | B   | C   |
| <b>Frame</b>     | 45 | ATT  | TTG | TGA | GAC | ACT | CTT | TTA | TAC | ACC | CCC | CCA | CAT | ATC | TCA | CAA |
| <b>Inj.1</b>     | 45 | B    | C   | B   | E   | A   | D   | C   | D   | E   | D   | C   | B   | F   | E   | D   |
| <b>Inj.2</b>     | 45 | B    | C   | B   | E   | B   | F   | C   | D   | E   | D   | C   | A   | D   | E   | D   |
| <b>Consensus</b> | 45 | B    | C   | B   | E   | *   | *   | C   | D   | E   | D   | C   | *   | *   | E   | D   |
| <b>Frame</b>     | 60 | AAC  | ACA | CAA | AAC | ACC | CCG | CGC | GCT | CTA | TAT | ATG | TGT | GTA | TAT | ATT |
| <b>Inj.1</b>     | 60 | C    | B   | C   | B   | E   | A   | D   | E   | B   | C   | A   | D   | F   | C   | B   |
| <b>Inj.2</b>     | 60 | C    | B   | C   | B   | E   | A   | D   | E   | B   | C   | A   | D   | F   | C   | B   |

Characterizing patterns in DNA sequence trace data through informatics tools

|                  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <b>Consensus</b> | 60  | C   | B   | C   | B   | E   | A   | D   | E   | B   | C   | A   | D   | F   | C   | B   |
| <b>Frame</b>     | 75  | TTT | TTC | TCG | CGT | GTA | TAC | ACA | CAT | ATT | TTA | TAC | ACT | CTG | TGC | GCC |
| <b>Inj.1</b>     | 75  | F   | F   | E   | A   | A   | A   | B   | F   | F   | C   | B   | E   | A   | D   | E   |
| <b>Inj.2</b>     | 75  | F   | F   | E   | A   | A   | A   | D   | E   | D   | C   | B   | E   | A   | D   | E   |
| <b>Consensus</b> | 75  | F   | F   | E   | A   | A   | A   | *   | *   | *   | C   | B   | E   | A   | D   | E   |
| <b>Frame</b>     | 90  | CCA | CAG | AGC | GCC | CCA | CAC | ACC | CCA | CAT | ATG | TGA | GAA | AAT | ATA | TAT |
| <b>Inj.1</b>     | 90  | D   | F   | C   | A   | B   | F   | E   | A   | D   | C   | B   | E   | D   | E   | D   |
| <b>Inj.2</b>     | 90  | D   | F   | C   | A   | B   | F   | E   | A   | D   | C   | B   | E   | D   | E   | D   |
| <b>Consensus</b> | 90  | D   | F   | C   | A   | B   | F   | E   | A   | D   | C   | B   | E   | D   | E   | D   |
| <b>Frame</b>     | 105 | ATT | TTG | TGT | GTA | TAC | ACG | CGG | GGT | GTA | TAC | ACC | CCA | CAT | ATA | TAA |
| <b>Inj.1</b>     | 105 | F   | C   | D   | E   | A   | B   | C   | D   | E   | A   | B   | F   | E   | A   | D   |
| <b>Inj.2</b>     | 105 | F   | C   | D   | E   | A   | B   | C   | D   | E   | A   | B   | F   | E   | A   | D   |
| <b>Consensus</b> | 105 | F   | C   | D   | E   | A   | B   | C   | D   | E   | A   | B   | F   | E   | A   | D   |
| <b>Frame</b>     | 120 | AAA | AAT | ATA | TAC | ACT | CTT | TTG | TGA | GAC | ACC | CCA | CAC | ACC | CCT | CTG |
| <b>Inj.1</b>     | 120 | C   | B   | E   | A   | D   | E   | A   | B   | F   | C   | A   | D   | F   | C   | A   |
| <b>Inj.2</b>     | 120 | C   | B   | E   | A   | D   | E   | A   | B   | F   | C   | A   | D   | F   | C   | A   |
| <b>Consensus</b> | 120 | C   | B   | E   | A   | D   | E   | A   | B   | F   | C   | A   | D   | F   | C   | A   |
| <b>Frame</b>     | 135 | TGT | GTA | TAG | AGT | GTA | TAC | ACA | CAT | ATA | TAA | AAA | --- | --- | AAC | ACC |
| <b>Inj.1</b>     | 135 | D   | F   | F   | C   | A   | A   | B   | F   | C   | D   | F   | -   | -   | C   | B   |
| <b>Inj.2</b>     | 135 | D   | F   | F   | C   | A   | A   | A   | D   | C   | D   | E   | -   | -   | C   | B   |
| <b>Consensus</b> | 135 | D   | F   | F   | C   | A   | A   | *   | *   | C   | D   | *   | -   | -   | C   | B   |
| <b>Frame</b>     | 150 | CCC | CCA | CAA | AAT | ATC | TCC | CCA | CAC | ACA | CAT | ATC | TCA | CAA | AAA | --- |
| <b>Inj.1</b>     | 150 | F   | E   | B   | E   | A   | B   | C   | D   | C   | B   | F   | E   | B   | C   | -   |
| <b>Inj.2</b>     | 150 | F   | E   | B   | E   | A   | B   | C   | D   | C   | B   | F   | E   | D   | C   | -   |
| <b>Consensus</b> | 150 | F   | E   | B   | E   | A   | B   | C   | D   | C   | B   | F   | E   | *   | C   | -   |
| <b>Frame</b>     | 165 | AAC | ACC | CCC | --- | --- | CCT | CTC | TCC | CCC | CCT | CTA | TAT | ATG | TGC | GCT |
| <b>Inj.1</b>     | 165 | E   | A   | D   | -   | -   | A   | B   | E   | D   | C   | A   | B   | C   | B   | F   |
| <b>Inj.2</b>     | 165 | E   | A   | D   | -   | -   | C   | B   | E   | D   | C   | A   | B   | C   | B   | F   |
| <b>Consensus</b> | 165 | E   | A   | D   | -   | -   | *   | B   | E   | D   | C   | A   | B   | C   | B   | F   |
| <b>Frame</b>     | 180 | CTT | TTA | TAC | ACA | CAA | AAG | AGC | GCA | CAA | AAG | AGT | GTA | TAC | ACA | CAG |
| <b>Inj.1</b>     | 180 | F   | C   | B   | F   | E   | B   | C   | D   | E   | B   | E   | D   | C   | B   | F   |
| <b>Inj.2</b>     | 180 | F   | C   | B   | F   | E   | D   | C   | D   | E   | B   | E   | D   | C   | B   | F   |
| <b>Consensus</b> | 180 | F   | C   | B   | F   | E   | *   | C   | D   | E   | B   | E   | D   | C   | B   | F   |
| <b>Frame</b>     | 195 | AGC | GCA | CAA | AAT | ATC | TCA | CAA | AAC | ACC | CCC | CCT | CTC | TCA | CAA | AAC |
| <b>Inj.1</b>     | 195 | F   | E   | B   | C   | B   | E   | A   | A   | A   | D   | E   | D   | E   | B   | C   |
| <b>Inj.2</b>     | 195 | F   | E   | D   | C   | B   | E   | A   | A   | A   | D   | E   | D   | E   | B   | E   |
| <b>Consensus</b> | 195 | F   | E   | *   | C   | B   | E   | A   | A   | A   | D   | E   | D   | E   | B   | *   |
| <b>Frame</b>     | 210 | ACT | CTA | TAT | ATC | TCA | CAC | ACA | CAC | ACA | CAT | ATC | TCA | CAA | AAC | ACT |
| <b>Inj.1</b>     | 210 | A   | B   | C   | D   | E   | B   | C   | A   | D   | E   | D   | E   | D   | C   | A   |
| <b>Inj.2</b>     | 210 | A   | B   | C   | D   | E   | B   | C   | A   | D   | E   | D   | E   | D   | C   | A   |
| <b>Consensus</b> | 210 | A   | B   | C   | D   | E   | B   | C   | A   | D   | E   | D   | E   | D   | C   | A   |
| <b>Frame</b>     | 225 | CTG | TGC | GCA | CAA | AAC | ACT | CTC | TCC | CCA | CAA | AAA | AAG | AGC | GCC | CCA |
| <b>Inj.1</b>     | 225 | A   | D   | F   | E   | A   | A   | D   | E   | D   | E   | A   | D   | C   | A   | D   |
| <b>Inj.2</b>     | 225 | A   | D   | F   | E   | A   | A   | D   | E   | D   | E   | A   | D   | C   | A   | D   |
| <b>Consensus</b> | 225 | A   | D   | F   | E   | A   | A   | D   | E   | D   | E   | A   | D   | C   | A   | D   |
| <b>Frame</b>     | 240 | CAC | ACC | CCC | --- | CCT | CTC | TCA | CAC | ACC | CCC | CCA | CAC | ACT | CTA | TAG |
| <b>Inj.1</b>     | 240 | E   | A   | B   | -   | C   | D   | E   | B   | C   | B   | F   | E   | A   | A   | D   |
| <b>Inj.2</b>     | 240 | E   | A   | B   | -   | C   | B   | E   | B   | C   | B   | F   | E   | A   | A   | D   |
| <b>Consensus</b> | 240 | E   | A   | B   | -   | C   | *   | E   | B   | C   | B   | F   | E   | A   | A   | D   |
| <b>Frame</b>     | 255 | AGG | GGA | GAT | ATA | TAT | ATC | TCA | CAA | AAC | ACA | CAA | AAA | AAC | ACC | CCT |
| <b>Inj.1</b>     | 255 | E   | A   | D   | E   | B   | F   | E   | B   | C   | D   | E   | A   | B   | F   | C   |
| <b>Inj.2</b>     | 255 | E   | A   | D   | E   | B   | F   | E   | B   | C   | D   | E   | A   | B   | F   | C   |
| <b>Consensus</b> | 255 | E   | A   | D   | E   | B   | F   | E   | B   | C   | D   | E   | A   | B   | F   | C   |
| <b>Frame</b>     | 270 | CTA | TAC | ACC | CCC | CCA | CAC | ACC | CCC | CCT | CTT | TTA | TAA | AAC | ACA | CAG |
| <b>Inj.1</b>     | 270 | D   | C   | B   | F   | E   | A   | A   | D   | E   | A   | A   | D   | E   | A   | D   |
| <b>Inj.2</b>     | 270 | D   | C   | B   | F   | E   | A   | A   | D   | E   | A   | B   | F   | E   | A   | D   |
| <b>Consensus</b> | 270 | D   | C   | B   | F   | E   | A   | A   | D   | E   | A   | *   | *   | E   | A   | D   |
| <b>Frame</b>     | 285 | AGT | GTA | TAC | ACA | CAT | ATA | TAG | AGT | GTA | TAC | ACA | CAT | ATA | TAA | AAA |
| <b>Inj.1</b>     | 285 | F   | F   | C   | B   | E   | A   | D   | E   | A   | A   | A   | B   | F   | F   | E   |
| <b>Inj.2</b>     | 285 | F   | F   | C   | B   | E   | A   | D   | E   | A   | A   | A   | B   | F   | F   | E   |
| <b>Consensus</b> | 285 | F   | F   | C   | B   | E   | A   | D   | E   | A   | A   | A   | B   | F   | F   | E   |
| <b>Frame</b>     | 300 | AAG | AGC | GCC | CCA | CAT | ATT | TTT | TTA | TAC | ACC | CCG | CGT | GTA | TAC | ACA |

|                  |     |     |     |     |     |     |     |      |     |     |     |     |     |     |     |      |
|------------------|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|------|
| <b>Inj.1</b>     | 300 | A   | A   | A   | D   | E   | B   | F    | E   | A   | B   | E   | D   | E   | A   | D    |
| <b>Inj.2</b>     | 300 | A   | A   | A   | D   | E   | B   | F    | E   | A   | D   | E   | D   | E   | A   | D    |
| <b>Consensus</b> | 300 | A   | A   | A   | D   | E   | B   | F    | E   | A   | *   | E   | D   | E   | A   | D    |
| <b>Frame</b>     | 315 | CAT | ATA | TAG | AGC | GCA | CAC | ACA  | CAT | ATT | TTA | TAC | ACA | CAG | AGT | GTC  |
| <b>Inj.1</b>     | 315 | E   | B   | F   | E   | A   | B   | C    | B   | F   | E   | B   | F   | F   | E   | A    |
| <b>Inj.2</b>     | 315 | E   | B   | F   | E   | A   | B   | C    | B   | F   | E   | A   | D   | F   | F   | C    |
| <b>Consensus</b> | 315 | E   | B   | F   | E   | A   | B   | C    | B   | F   | E   | *   | *   | F   | *   | *    |
| <b>Frame</b>     | 330 | TCA | CAA | AAA | AAT | ATC | TCC | CCC  | CCT | CTT | TTC | TCT | CTC | TCG | CGC | GCC  |
| <b>Inj.1</b>     | 330 | D   | C   | B   | E   | B   | C   | D    | C   | A   | D   | E   | D   | E   | A   | A    |
| <b>Inj.2</b>     | 330 | D   | C   | B   | C   | B   | C   | D    | C   | A   | D   | E   | D   | E   | A   | A    |
| <b>Consensus</b> | 330 | D   | C   | B   | *   | B   | C   | D    | C   | A   | D   | E   | D   | E   | A   | A    |
| <b>Frame</b>     | 345 | CCC | --- | --- | CCA | CAT | ATG | TGG  | GGA | GAT | ATG | TGA | GAC | ACC | CCC | ---  |
| <b>Inj.1</b>     | 345 | B   | -   | -   | D   | C   | A   | D    | E   | D   | C   | A   | D   | E   | D   | -    |
| <b>Inj.2</b>     | 345 | B   | -   | -   | D   | C   | A   | BETA | E   | D   | C   | A   | D   | E   | D   | -    |
| <b>Consensus</b> | 345 | B   | -   | -   | D   | C   | A   | *    | E   | D   | C   | A   | D   | E   | D   | -    |
| <b>Frame</b>     | 360 | --- | --- | CCT | CTC | TCA | CAG | AGA  | GAT | ATA | TAG | AGG | GGG | --- | GGT | GTC  |
| <b>Inj.1</b>     | 360 | -   | -   | C   | D   | E   | D   | C    | B   | E   | D   | C   | B   | -   | F   | ALFA |
| <b>Inj.2</b>     | 360 | -   | -   | A   | D   | E   | D   | C    | B   | E   | D   | C   | B   | -   | F   | F    |
| <b>Consensus</b> | 360 | -   | -   | *   | D   | E   | D   | C    | B   | E   | D   | C   | B   | -   | F   | *    |
| <b>Frame</b>     | 375 | TCC | CCC | CCT | CTT | TTG |     |      |     |     |     |     |     |     |     |      |
| <b>Inj.1</b>     | 375 | PSI | D   | E   | A   | A   |     |      |     |     |     |     |     |     |     |      |
| <b>Inj.2</b>     | 375 | C   | D   | E   | A   | A   |     |      |     |     |     |     |     |     |     |      |
| <b>Consensus</b> | 375 | *   | D   | E   | A   | A   |     |      |     |     |     |     |     |     |     |      |

**Table B.** Two different samples sequenced with the A1 primer electrophoresed on the same instrument in the same run. The frames displayed in this example are from p16,043 to p16,186.

|                  |     |     |      |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------------------|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <b>Frame</b>     | 1   | ATT | TTT  | TTG | TGG | GGG | GGT | GTA | TAC | ACC | CCA | CAC | ACC | CCC | CCA |     |
| <b>Sample 1</b>  | 1   | F   | ALFA | PSI | B   | F   | F   | C   | A   | D   | E   | D   | C   | B   | E   |     |
| <b>Sample 2</b>  | 1   | F   | F    | C   | B   | F   | F   | C   | A   | D   | E   | D   | C   | B   | C   |     |
| <b>Consensus</b> | 1   | F   | *    | *   | B   | F   | F   | C   | A   | D   | E   | D   | C   | B   | *   |     |
| <b>Frame</b>     | 15  | CAA | AAG  | AGT | GTA | TAT | ATT | TTG | TGA | GAC | ACT | CTC | TCA | CAC | ACC | CCC |
| <b>Sample 1</b>  | 15  | D   | E    | A   | B   | C   | B   | C   | B   | E   | A   | D   | E   | B   | C   | B   |
| <b>Sample 2</b>  | 15  | D   | F    | C   | B   | C   | D   | C   | B   | E   | B   | F   | E   | B   | E   | B   |
| <b>Consensus</b> | 15  | D   | *    | *   | B   | C   | *   | C   | B   | E   | *   | *   | E   | B   | *   | B   |
| <b>Frame</b>     | 30  | CCA | CAT  | ATC | TCA | CAA | AAC | ACA | CAA | AAC | ACC | CCG | CGC | GCT | CTA | TAT |
| <b>Sample 1</b>  | 30  | E   | A    | D   | C   | B   | E   | B   | C   | B   | E   | A   | D   | E   | B   | C   |
| <b>Sample 2</b>  | 30  | C   | A    | D   | C   | B   | E   | B   | C   | D   | E   | A   | D   | E   | A   | B   |
| <b>Consensus</b> | 30  | *   | A    | D   | C   | B   | E   | B   | C   | *   | E   | A   | D   | E   | *   | #   |
| <b>Frame</b>     | 45  | ATG | TGT  | GTA | TAT | ATT | TTT | TTC | TCG | CGT | GTA | TAC | ACA | CAT | ATT | TTA |
| <b>Sample 1</b>  | 45  | A   | D    | E   | A   | D   | F   | F   | E   | A   | A   | A   | B   | F   | F   | C   |
| <b>Sample 2</b>  | 45  | C   | D    | E   | A   | D   | F   | F   | E   | A   | A   | A   | D   | F   | F   | C   |
| <b>Consensus</b> | 45  | *   | D    | E   | A   | D   | F   | F   | E   | A   | A   | A   | *   | F   | F   | C   |
| <b>Frame</b>     | 60  | TAC | ACT  | CTG | TGC | GCC | CCA | CAG | AGC | GCC | CCA | CAC | ACC | CCA | CAT | ATG |
| <b>Sample 1</b>  | 60  | B   | E    | A   | D   | E   | D   | F   | C   | A   | B   | F   | E   | A   | D   | C   |
| <b>Sample 2</b>  | 60  | B   | E    | A   | D   | E   | D   | F   | C   | A   | B   | F   | F   | C   | D   | C   |
| <b>Consensus</b> | 60  | B   | E    | A   | D   | E   | D   | F   | C   | A   | B   | F   | *   | *   | D   | C   |
| <b>Frame</b>     | 75  | TGA | GAA  | AAT | ATA | TAT | ATT | TTG | -   | -   | -   | ACG | CGG | GGT | GTA | TAC |
| <b>Sample 1</b>  | 75  | B   | E    | D   | E   | D   | F   | C   | D   | E   | A   | B   | C   | D   | E   | A   |
| <b>Sample 2</b>  | 75  | B   | E    | D   | E   | D   | F   | C   | D   | E   | A   | D   | C   | D   | E   | A   |
| <b>Consensus</b> | 75  | B   | E    | D   | E   | D   | F   | C   | *   | *   | *   | *   | C   | D   | E   | A   |
| <b>Frame</b>     | 90  | ACC | CCA  | CAT | ATA | TAA | AAA | AAT | ATA | TAC | ACT | CTT | TTG | TGA | GAC | ACC |
| <b>Sample 1</b>  | 90  | B   | F    | C   | A   | D   | C   | B   | E   | A   | D   | E   | A   | D   | F   | C   |
| <b>Sample 2</b>  | 90  | B   | F    | E   | A   | D   | C   | B   | E   | A   | D   | E   | A   | D   | F   | C   |
| <b>Consensus</b> | 90  | B   | F    | *   | A   | D   | C   | B   | E   | A   | D   | E   | A   | D   | F   | C   |
| <b>Frame</b>     | 105 | CCA | CAC  | ACC | CCT | CTG | TGT | GTA | TAG | AGT | GTA | TAC | ACA | CAT | ATA | TAA |
| <b>Sample 1</b>  | 105 | A   | D    | F   | C   | A   | D   | F   | F   | C   | A   | A   | B   | F   | C   | D   |
| <b>Sample 2</b>  | 105 | A   | D    | F   | C   | A   | D   | F   | F   | C   | A   | A   | A   | D   | C   | D   |
| <b>Consensus</b> | 105 | A   | D    | F   | C   | A   | D   | F   | F   | C   | A   | A   | *   | *   | C   | D   |
| <b>Frame</b>     | 120 | AAA | ---  | --- | AAC | ACC | CCC | CCA | CAA | AAT | ATC | TCC | CCA | CAC | ACA | CAT |
| <b>Sample 1</b>  | 120 | E   | -    | -   | C   | B   | F   | E   | D   | E   | A   | B   | C   | D   | E   | A   |
| <b>Sample 2</b>  | 120 | E   | -    | -   | C   | B   | F   | E   | B   | F   | C   | B   | C   | D   | E   | B   |
| <b>Consensus</b> | 120 | E   | -    | -   | C   | B   | F   | E   | *   | *   | *   | B   | C   | D   | E   | *   |
| <b>Frame</b>     | 135 | ATC | TCA  | CAA | AAA | --- | AAC | ACC | CCC | -   | -   | -   | -   | -   | -   | -   |
| <b>Sample 1</b>  | 135 | D   | E    | D   | C   | -   | C   | A   | D   | -   | -   | A   | B   | E   | -   | -   |
| <b>Sample 2</b>  | 135 | F   | E    | B   | C   | -   | E   | A   | D   | E   | B   | E   | D   | -   | -   | -   |
| <b>Consensus</b> | 135 | *   | E    | *   | C   | -   | *   | A   | D   | *   | *   | #   | *   | *   | -   | -   |

**Table C.** The same sample sequenced with the B1 primer electrophoresed on two different instruments in same laboratory. The frames displayed in this example are from p16,367 to p16,034.

|                  |     |      |       |     |     |     |     |     |     |     |     |     |      |     |     |     |
|------------------|-----|------|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|------|-----|-----|-----|
| <b>Frame</b>     | 1   | -    | -     | -   | -   | -   | -   | -   | -   | -   | --- | --- | -    | -   | -   | -   |
| <b>Inst. 1</b>   | 1   | -    | -     | -   | -   | -   | -   | -   | -   | -   | -   | -   | -    | -   | -   | -   |
| <b>Inst. 2</b>   | 1   | A    | B     | F   | C   | A   | B   | F   | E   | -   | -   | D   | C    | B   | F   |     |
| <b>Consensus</b> | 1   | *    | *     | *   | *   | *   | *   | *   | *   | *   | -   | -   | *    | *   | *   | *   |
| <b>Frame</b>     | 15  | -    | -     | -   | -   | TGG | GGG | --- | GGA | GAC | ACG | CGA | GAG  | AGA | GAA | AAG |
| <b>Inst. 1</b>   | 15  | -    | -     | -   | -   | B   | F   | -   | A   | A   | D   | E   | D    | E   | B   | E   |
| <b>Inst. 2</b>   | 15  | E    | A     | A   | B   | F   | F   | -   | A   | A   | D   | E   | D    | E   | B   | E   |
| <b>Consensus</b> | 15  | *    | *     | *   | *   | #   | F   | -   | A   | A   | D   | E   | D    | E   | B   | E   |
| <b>Frame</b>     | 30  | AGG  | GGG   | GGA | GAT | ATT | TTT | TTG | TGA | GAC | ACT | CTG | TGT  | GTA | TAA | AAT |
| <b>Inst. 1</b>   | 30  | A    | D     | E   | B   | F   | F   | E   | A   | A   | A   | A   | D    | F   | E   | A   |
| <b>Inst. 2</b>   | 30  | A    | D     | E   | A   | D   | F   | E   | A   | A   | A   | A   | D    | F   | E   | A   |
| <b>Consensus</b> | 30  | A    | D     | E   | *   | *   | F   | E   | A   | A   | A   | A   | D    | F   | E   | A   |
| <b>Frame</b>     | 45  | ATG  | TGT   | GTG | TGC | GCT | CTA | TAT | ATG | TGT | GTA | TAC | ACG  | CGA | GAT | ATG |
| <b>Inst. 1</b>   | 45  | A    | D     | E   | A   | D   | F   | C   | A   | D   | F   | C   | B    | C   | B   | E   |
| <b>Inst. 2</b>   | 45  | A    | D     | E   | A   | D   | F   | C   | A   | D   | F   | C   | B    | C   | B   | E   |
| <b>Consensus</b> | 45  | A    | D     | E   | A   | D   | F   | C   | A   | D   | F   | C   | B    | C   | B   | E   |
| <b>Frame</b>     | 60  | TGA  | GAA   | AAT | ATG | TGG | GGC | GCT | CTT | TTT | TTA | TAT | ATG  | TGT | GTA | TAC |
| <b>Inst. 1</b>   | 60  | A    | A     | D   | C   | B   | F   | E   | B   | F   | F   | C   | B    | F   | F   | C   |
| <b>Inst. 2</b>   | 60  | A    | A     | D   | C   | D   | F   | E   | B   | F   | F   | C   | A    | B   | F   | C   |
| <b>Consensus</b> | 60  | A    | A     | D   | C   | *   | F   | E   | B   | F   | F   | C   | *    | #   | F   | C   |
| <b>Frame</b>     | 75  | ACT  | CTA   | TAT | ATG | TGT | GTA | TAC | ACT | CTG | TGT | GTT | TTG  | TGA | GAG | AGG |
| <b>Inst. 1</b>   | 75  | B    | E     | B   | E   | D   | F   | C   | B   | C   | D   | E   | D    | E   | D   | C   |
| <b>Inst. 2</b>   | 75  | B    | E     | B   | E   | D   | F   | C   | B   | C   | D   | E   | D    | E   | D   | C   |
| <b>Consensus</b> | 75  | B    | E     | B   | E   | D   | F   | C   | B   | C   | D   | E   | D    | E   | D   | C   |
| <b>Frame</b>     | 90  | GGA  | GAT   | ATG | TGG | GGG | GGT | GTA | TAG | AGG | GGT | GTT | TTT  | TTG | TGT | GTT |
| <b>Inst. 1</b>   | 90  | B    | F     | C   | B   | F   | E   | B   | F   | C   | B   | F   | F    | C   | D   | E   |
| <b>Inst. 2</b>   | 90  | B    | F     | C   | B   | F   | E   | D   | F   | C   | B   | F   | F    | C   | D   | E   |
| <b>Consensus</b> | 90  | B    | F     | C   | B   | F   | E   | *   | F   | C   | B   | F   | F    | C   | D   | E   |
| <b>Frame</b>     | 105 | TTG  | TGG   | GGT | GTA | TAT | ATC | TCC | CCT | CTA | TAG | AGT | GTG  | TGG | GGG | GGT |
| <b>Inst. 1</b>   | 105 | A    | B     | F   | F   | C   | B   | E   | D   | C   | D   | E   | A    | D   | E   | B   |
| <b>Inst. 2</b>   | 105 | A    | B     | F   | F   | C   | B   | E   | B   | C   | D   | E   | A    | D   | E   | B   |
| <b>Consensus</b> | 105 | A    | B     | F   | F   | C   | B   | E   | *   | C   | D   | E   | A    | D   | E   | B   |
| <b>Frame</b>     | 120 | GTG  | TGA   | GAG | AGG | GGG | --- | GGT | GTG | TGG | GCC | GCT | CTT  | TTT | TTG | TGG |
| <b>Inst. 1</b>   | 120 | C    | A     | D   | C   | B   | -   | D   | C   | B   | F   | C   | B    | F   | C   | D   |
| <b>Inst. 2</b>   | 120 | C    | A     | D   | C   | B   | -   | D   | C   | B   | F   | C   | B    | F   | C   | D   |
| <b>Consensus</b> | 120 | C    | A     | D   | C   | B   | -   | D   | C   | B   | F   | C   | B    | F   | C   | D   |
| <b>Frame</b>     | 135 | GGA  | GAG   | AGT | GTT | TTG | TGC | GCA | CAG | AGT | GTT | TTG | TGA  | GAT | ATG | TGT |
| <b>Inst. 1</b>   | 135 | E    | B     | C   | B   | F   | F   | F   | F   | E   | A   | B   | C    | D   | C   | B   |
| <b>Inst. 2</b>   | 135 | E    | B     | C   | B   | F   | F   | F   | F   | E   | A   | A   | A    | D   | C   | B   |
| <b>Consensus</b> | 135 | E    | B     | C   | B   | F   | F   | F   | F   | E   | A   | *   | *    | D   | C   | B   |
| <b>Frame</b>     | 150 | GTG  | TGT   | GTG | TGA | GAT | ATA | TAG | AGT | GTT | TTG | TGA | GAA  | AAG | AGG | GGT |
| <b>Inst. 1</b>   | 150 | E    | D     | C   | A   | D   | E   | D   | E   | A   | B   | C   | B    | F   | C   | B   |
| <b>Inst. 2</b>   | 150 | E    | D     | C   | A   | D   | E   | D   | E   | A   | B   | C   | BETA | F   | C   | B   |
| <b>Consensus</b> | 150 | E    | D     | C   | A   | D   | E   | D   | E   | A   | B   | C   | *    | F   | C   | B   |
| <b>Frame</b>     | 165 | GTT  | TTG   | TGA | GAT | ATT | TTG | TGC | GCT | CTG | TGT | GTA | TAC  | ACT | CTT | TTG |
| <b>Inst. 1</b>   | 165 | F    | C     | A   | D   | E   | A   | D   | E   | A   | D   | F   | C    | A   | A   | A   |
| <b>Inst. 2</b>   | 165 | F    | C     | A   | D   | E   | A   | D   | E   | A   | D   | F   | C    | A   | A   | A   |
| <b>Consensus</b> | 165 | F    | C     | A   | D   | E   | A   | D   | E   | A   | D   | F   | C    | A   | A   | A   |
| <b>Frame</b>     | 180 | TGC  | GCT   | CTT | TTG | TGT | GTA | TAA | AAG | AGC | GCA | CAT | ATG  | TGG | GGG | --- |
| <b>Inst. 1</b>   | 180 | D    | E     | B   | C   | D   | F   | C   | B   | C   | B   | E   | A    | D   | F   | -   |
| <b>Inst. 2</b>   | 180 | D    | E     | B   | C   | D   | F   | C   | B   | C   | B   | E   | A    | D   | F   | -   |
| <b>Consensus</b> | 180 | D    | E     | B   | C   | D   | F   | C   | B   | C   | B   | E   | A    | D   | F   | -   |
| <b>Frame</b>     | 195 | GGA  | GAG   | AGG | GGG | --- | --- | GGT | GTT | TTT | --- | TTG | TGA  | GAT | ATG | TGT |
| <b>Inst. 1</b>   | 195 | A    | B     | C   | B   | -   | -   | D   | F   | F   | -   | C   | A    | D   | C   | B   |
| <b>Inst. 2</b>   | 195 | A    | A     | A   | B   | -   | -   | D   | F   | F   | -   | C   | B    | F   | C   | B   |
| <b>Consensus</b> | 195 | A    | *     | *   | B   | -   | -   | D   | F   | F   | -   | C   | *    | *   | C   | B   |
| <b>Frame</b>     | 210 | GTG  | TGG   | GGA | GAT | ATT | TTG | TGG | GGG | GGT | GTT | TTT | ---  | --- | TTA | TAT |
| <b>Inst. 1</b>   | 210 | ALFA | GAMMA | C   | B   | F   | C   | D   | E   | A   | A   | D   | -    | -   | C   | B   |
| <b>Inst. 2</b>   | 210 | E    | D     | C   | B   | F   | C   | D   | E   | A   | A   | D   | -    | -   | C   | B   |

Characterizing patterns in DNA sequence trace data through informatics tools

|                  |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| <b>Consensus</b> | 210 | *   | #   | C   | B   | F   | C   | D   | E   | A   | A   | D   | -   | -   | C   | B   |
| <b>Frame</b>     | 225 | ATG | TGT | GTA | TAC | ACT | CTA | TAC | ACA | CAG | AGG | GGT | GTG | TGG | GGT | GTC |
| <b>Inst. 1</b>   | 225 | F   | F   | F   | C   | A   | A   | D   | F   | F   | C   | B   | C   | B   | F   | E   |
| <b>Inst. 2</b>   | 225 | F   | F   | F   | C   | B   | C   | D   | F   | F   | C   | B   | C   | B   | F   | E   |
| <b>Consensus</b> | 225 | F   | F   | F   | C   | *   | *   | D   | F   | F   | C   | B   | C   | B   | F   | E   |
| <b>Frame</b>     | 240 | TCA | CAA | AAG | AGT | GTA | TAT | ATT | TTT | TTA | TAT | ATG | TGG | GGT | GTA | TAC |
| <b>Inst. 1</b>   | 240 | A   | A   | D   | C   | B   | C   | B   | F   | C   | A   | D   | E   | D   | E   | A   |
| <b>Inst. 2</b>   | 240 | A   | A   | D   | C   | B   | E   | B   | F   | C   | A   | D   | E   | D   | E   | A   |
| <b>Consensus</b> | 240 | A   | A   | D   | C   | B   | *   | B   | F   | C   | A   | D   | E   | D   | E   | A   |
| <b>Frame</b>     | 255 | ACC | CCG | CGT | GTA | TAC | ACA | CAA | AAT | ATA | TAT | ATT | TTC | TCA | CAT | ATG |
| <b>Inst. 1</b>   | 255 | B   | F   | E   | B   | C   | B   | F   | C   | A   | B   | F   | F   | E   | A   | A   |
| <b>Inst. 2</b>   | 255 | B   | F   | E   | B   | C   | B   | F   | E   | A   | B   | F   | F   | E   | A   | A   |
| <b>Consensus</b> | 255 | B   | F   | E   | B   | C   | B   | F   | *   | A   | B   | F   | F   | E   | A   | A   |
| <b>Frame</b>     | 270 | TGG | GGT | GTG | TGG | GGC | GCT | CTG | TGG | GGC | GCA | CAG | AGT | GTA | TAA | AAT |
| <b>Inst. 1</b>   | 270 | D   | F   | C   | B   | F   | C   | A   | D   | F   | E   | D   | E   | B   | C   | A   |
| <b>Inst. 2</b>   | 270 | B   | F   | C   | B   | F   | C   | A   | D   | F   | F   | F   | C   | B   | C   | A   |
| <b>Consensus</b> | 270 | *   | F   | C   | B   | F   | C   | A   | D   | F   | *   | *   | *   | B   | C   | A   |
| <b>Frame</b>     | 285 | ATG | TGT | GTA | TAC | ACG | CGA | GAA | AAA | AAT | ATA | TAC | ACA | CAT | ATA | TAG |
| <b>Inst. 1</b>   | 285 | A   | D   | E   | A   | D   | C   | B   | E   | B   | C   | B   | F   | C   | B   | F   |
| <b>Inst. 2</b>   | 285 | A   | D   | E   | A   | D   | C   | B   | E   | B   | C   | B   | F   | C   | B   | F   |
| <b>Consensus</b> | 285 | A   | D   | E   | A   | D   | C   | B   | E   | B   | C   | B   | F   | C   | B   | F   |
| <b>Frame</b>     | 300 | AGC | GCG | CGG | GGT | GTT | TTG | TGT | GTT | TTG | TGA | GAT | ATG | TGG | GGG | GGT |
| <b>Inst. 1</b>   | 300 | E   | A   | A   | D   | C   | A   | D   | E   | A   | A   | D   | E   | B   | F   | E   |
| <b>Inst. 2</b>   | 300 | E   | A   | B   | F   | C   | A   | D   | E   | A   | A   | D   | C   | B   | F   | E   |
| <b>Consensus</b> | 300 | E   | A   | *   | *   | C   | A   | D   | E   | A   | A   | D   | *   | B   | F   | E   |
| <b>Frame</b>     | 315 | GTG | TGA | GAG | AGT | GTC | TCA | CAA | AAT | ATA | TAC | ACT | CTT | TTG | TGG | GGG |
| <b>Inst. 1</b>   | 315 | A   | A   | D   | E   | B   | C   | B   | E   | D   | E   | A   | A   | D   | F   | E   |
| <b>Inst. 2</b>   | 315 | A   | A   | D   | E   | B   | C   | B   | E   | B   | E   | A   | A   | D   | F   | E   |
| <b>Consensus</b> | 315 | A   | A   | D   | E   | B   | C   | B   | E   | *   | E   | A   | A   | D   | F   | E   |
| <b>Frame</b>     | 330 | GGT | GTG | TGG | GGT | GTA | TAC | ACC | CCC | CCA | CAA | AAA | AAT | ATC | TCT | CTG |
| <b>Inst. 1</b>   | 330 | A   | A   | B   | F   | E   | A   | B   | F   | F   | F   | C   | A   | D   | C   | A   |
| <b>Inst. 2</b>   | 330 | A   | A   | B   | F   | CHI | A   | B   | F   | F   | F   | C   | A   | D   | C   | A   |
| <b>Consensus</b> | 330 | A   | A   | B   | F   | *   | A   | B   | F   | F   | F   | C   | A   | D   | C   | A   |
| <b>Frame</b>     | 345 | TGC | GCT | CTT | TTC | TCC | CCC |     |     |     |     |     |     |     |     |     |
| <b>Inst. 1</b>   | 345 | D   | E   | A   | D   | E   | B   |     |     |     |     |     |     |     |     |     |
| <b>Inst. 2</b>   | 345 | D   | E   | B   | F   | E   | D   |     |     |     |     |     |     |     |     |     |
| <b>Consensus</b> | 345 | D   | E   | *   | *   | E   | *   |     |     |     |     |     |     |     |     |     |



## Appendix II — Frame Distributions Within a Laboratory

The following tables are examples of the processed data using the bioinformatics tools designed to statistically evaluate each of the 64 frames using Chi Square analysis. Each of the four primers, A1, B1, C1, and D1, were evaluated to characterize the different patterns for each frame for two different laboratories. The white cells are the observed number of occurrences of each pattern. The peach cells are the expected number of occurrences under the null hypothesis that each of the six patterns is equally likely. The Chi Square contribution for each pattern is in the corresponding purple cell. The totals correspond to the total number of observations and the total Chi Square. The p-value is the probability of observing a Chi Square greater than or equal to that observed under the null hypothesis of a uniform distribution with each cell having a relative frequency of 1/6 with 5 degrees of freedom.

**Table A.** Laboratory A data compiled for results obtained from Primer A1.

|                | A      | B      | C      | D      | E      | F      | TOTALS  |             |
|----------------|--------|--------|--------|--------|--------|--------|---------|-------------|
| AAA            | 199.00 | 87.00  | 264.00 | 1.00   | 159.00 | 85.00  | 795.00  |             |
| Exp.           | 132.50 | 132.50 | 132.50 | 132.50 | 132.50 | 132.50 |         | p-value     |
| X <sup>2</sup> | 33.38  | 15.62  | 130.51 | 130.51 | 5.30   | 17.03  | 332.34  | 1.10576E-69 |
| AAC            | 296.00 | 271.00 | 353.00 | 26.00  | 217.00 | 75.00  | 1238.00 |             |
| Exp.           | 206.33 | 206.33 | 206.33 | 206.33 | 206.33 | 206.33 |         | p-value     |
| X <sup>2</sup> | 38.97  | 20.27  | 104.25 | 157.61 | 0.55   | 83.60  | 405.24  | 2.19786E-85 |
| ACA            | 174.00 | 330.00 | 414.00 | 380.00 | 56.00  | 204.00 | 1558.00 |             |
| Exp.           | 259.67 | 259.67 | 259.67 | 259.67 | 259.67 | 259.67 |         | p-value     |
| X <sup>2</sup> | 28.26  | 19.05  | 91.73  | 55.76  | 159.74 | 11.93  | 366.48  | 4.94025E-77 |
| ACC            | 269.00 | 503.00 | 386.00 | 318.00 | 288.00 | 268.00 | 2032.00 |             |
| Exp.           | 338.67 | 338.67 | 338.67 | 338.67 | 338.67 | 338.67 |         | p-value     |
| X <sup>2</sup> | 14.33  | 79.74  | 6.62   | 1.26   | 7.58   | 14.75  | 124.27  | 3.90081E-25 |
| ACT            | 445.00 | 58.00  | 98.00  | 104.00 | 112.00 | 2.00   | 819.00  |             |
| Exp.           | 136.50 | 136.50 | 136.50 | 136.50 | 136.50 | 136.50 |         | p-value     |
| X <sup>2</sup> | 697.23 | 45.14  | 10.86  | 7.74   | 4.40   | 132.53 | 897.90  | 7.5747E-192 |
| ACG            | 1.00   | 98.00  | 0.00   | 11.00  | 0.00   | 1.00   | 111.00  |             |
| Exp.           | 18.50  | 18.50  | 18.50  | 18.50  | 18.50  | 18.50  |         | p-value     |
| X <sup>2</sup> | 16.55  | 341.64 | 18.50  | 3.04   | 18.50  | 16.55  | 414.78  | 1.9299E-87  |
| AAT            | 13.00  | 120.00 | 81.00  | 66.00  | 162.00 | 110.00 | 552.00  |             |
| Exp.           | 92.00  | 92.00  | 92.00  | 92.00  | 92.00  | 92.00  |         | p-value     |
| X <sup>2</sup> | 67.84  | 8.52   | 1.32   | 7.35   | 53.26  | 3.52   | 141.80  | 7.39751E-29 |
| ATA            | 185.00 | 94.00  | 148.00 | 5.00   | 445.00 | 18.00  | 895.00  |             |
| Exp.           | 149.17 | 149.17 | 149.17 | 149.17 | 149.17 | 149.17 |         | p-value     |
| X <sup>2</sup> | 8.61   | 20.40  | 0.01   | 139.33 | 586.71 | 115.34 | 870.40  | 6.7716E-186 |
| ATT            | 249.00 | 199.00 | 26.00  | 70.00  | 238.00 | 82.00  | 864.00  |             |
| Exp.           | 144.00 | 144.00 | 144.00 | 144.00 | 144.00 | 144.00 |         | p-value     |
| X <sup>2</sup> | 76.56  | 21.01  | 96.69  | 38.03  | 61.36  | 26.69  | 320.35  | 4.21498E-67 |
| ATC            | 93.00  | 222.00 | 53.00  | 231.00 | 13.00  | 170.00 | 782.00  |             |

Characterizing patterns in DNA sequence trace data through informatics tools

|      |         |        |        |         |        |        |                |                |
|------|---------|--------|--------|---------|--------|--------|----------------|----------------|
| Exp. | 130.33  | 130.33 | 130.33 | 130.33  | 130.33 | 130.33 |                | <b>p-value</b> |
| X^2  | 10.69   | 64.47  | 45.89  | 77.75   | 105.63 | 12.07  | <b>316.51</b>  | 2.82496E-66    |
| ATG  | 193.00  | 0.00   | 267.00 | 0.00    | 3.00   | 0.00   | 463.00         |                |
| Exp. | 77.17   | 77.17  | 77.17  | 77.17   | 77.17  | 77.17  |                | <b>p-value</b> |
| X^2  | 173.88  | 77.17  | 467.00 | 77.17   | 71.28  | 77.17  | <b>943.66</b>  | 9.4623E-202    |
| AAG  | 73.00   | 176.00 | 26.00  | 219.00  | 54.00  | 61.00  | 609.00         |                |
| Exp. | 101.50  | 101.50 | 101.50 | 101.50  | 101.50 | 101.50 |                | <b>p-value</b> |
| X^2  | 8.00    | 54.68  | 56.16  | 136.02  | 22.23  | 16.16  | <b>293.26</b>  | 2.82071E-61    |
| AGA  | 1.00    | 0.00   | 143.00 | 0.00    | 0.00   | 0.00   | 144.00         |                |
| Exp. | 24.00   | 24.00  | 24.00  | 24.00   | 24.00  | 24.00  |                | <b>p-value</b> |
| X^2  | 22.04   | 24.00  | 590.04 | 24.00   | 24.00  | 24.00  | <b>708.08</b>  | 8.7789E-151    |
| AGG  | 0.00    | 0.00   | 46.00  | 0.00    | 101.00 | 0.00   | 147.00         |                |
| Exp. | 24.50   | 24.50  | 24.50  | 24.50   | 24.50  | 24.50  |                | <b>p-value</b> |
| X^2  | 24.50   | 24.50  | 18.87  | 24.50   | 238.87 | 24.50  | <b>355.73</b>  | 1.01943E-74    |
| AGC  | 96.00   | 32.00  | 281.00 | 17.00   | 240.00 | 92.00  | 758.00         |                |
| Exp. | 126.33  | 126.33 | 126.33 | 126.33  | 126.33 | 126.33 |                | <b>p-value</b> |
| X^2  | 7.28    | 70.44  | 189.35 | 94.62   | 102.27 | 9.33   | <b>473.30</b>  | 4.6226E-100    |
| AGT  | 75.00   | 0.00   | 186.00 | 1.00    | 305.00 | 9.00   | 576.00         |                |
| Exp. | 96.00   | 96.00  | 96.00  | 96.00   | 96.00  | 96.00  |                | <b>p-value</b> |
| X^2  | 4.59    | 96.00  | 84.38  | 94.01   | 455.01 | 78.84  | <b>812.83</b>  | 1.936E-173     |
| CAA  | 75.00   | 261.00 | 216.00 | 427.00  | 617.00 | 81.00  | 1677.00        |                |
| Exp. | 279.50  | 279.50 | 279.50 | 279.50  | 279.50 | 279.50 |                | <b>p-value</b> |
| X^2  | 149.63  | 1.22   | 14.43  | 77.84   | 407.54 | 140.97 | <b>791.63</b>  | 7.4958E-169    |
| CAC  | 126.00  | 481.00 | 30.00  | 270.00  | 213.00 | 222.00 | 1342.00        |                |
| Exp. | 223.67  | 223.67 | 223.67 | 223.67  | 223.67 | 223.67 |                | <b>p-value</b> |
| X^2  | 42.65   | 296.07 | 167.69 | 9.60    | 0.51   | 0.01   | <b>516.52</b>  | 2.1631E-109    |
| CCA  | 145.00  | 120.00 | 313.00 | 422.00  | 276.00 | 716.00 | 1992.00        |                |
| Exp. | 332.00  | 332.00 | 332.00 | 332.00  | 332.00 | 332.00 |                | <b>p-value</b> |
| X^2  | 105.33  | 135.37 | 1.09   | 24.40   | 9.45   | 444.14 | <b>719.78</b>  | 2.599E-153     |
| CCC  | 18.00   | 366.00 | 7.00   | 520.00  | 8.00   | 378.00 | 1297.00        |                |
| Exp. | 216.17  | 216.17 | 216.17 | 216.17  | 216.17 | 216.17 |                | <b>p-value</b> |
| X^2  | 181.67  | 103.86 | 202.39 | 427.05  | 200.46 | 121.16 | <b>1236.59</b> | 3.4899E-265    |
| CCT  | 98.00   | 29.00  | 312.00 | 37.00   | 275.00 | 36.00  | 787.00         |                |
| Exp. | 131.17  | 131.17 | 131.17 | 131.17  | 131.17 | 131.17 |                | <b>p-value</b> |
| X^2  | 8.39    | 79.58  | 249.31 | 67.60   | 157.72 | 69.05  | <b>631.65</b>  | 2.9344E-134    |
| CCG  | 93.00   | 12.00  | 25.00  | 4.00    | 78.00  | 19.00  | 231.00         |                |
| Exp. | 38.50   | 38.50  | 38.50  | 38.50   | 38.50  | 38.50  |                | <b>p-value</b> |
| X^2  | 77.15   | 18.24  | 4.73   | 30.92   | 40.53  | 9.88   | <b>181.44</b>  | 2.63361E-37    |
| CAT  | 43.00   | 368.00 | 217.00 | 245.00  | 555.00 | 71.00  | 1499.00        |                |
| Exp. | 249.83  | 249.83 | 249.83 | 249.83  | 249.83 | 249.83 |                | <b>p-value</b> |
| X^2  | 171.23  | 55.89  | 4.31   | 0.09    | 372.76 | 128.01 | <b>732.30</b>  | 5.091E-156     |
| CTA  | 212.00  | 194.00 | 4.00   | 55.00   | 3.00   | 47.00  | 515.00         |                |
| Exp. | 85.83   | 85.83  | 85.83  | 85.83   | 85.83  | 85.83  |                | <b>p-value</b> |
| X^2  | 185.45  | 136.31 | 78.02  | 11.08   | 79.94  | 17.57  | <b>508.37</b>  | 1.2478E-107    |
| CTT  | 282.00  | 11.00  | 19.00  | 0.00    | 145.00 | 76.00  | 533.00         |                |
| Exp. | 88.83   | 88.83  | 88.83  | 88.83   | 88.83  | 88.83  |                | <b>p-value</b> |
| X^2  | 420.04  | 68.20  | 54.90  | 88.83   | 35.51  | 1.85   | <b>669.33</b>  | 2.099E-142     |
| CTC  | 1.00    | 175.00 | 2.00   | 528.00  | 3.00   | 46.00  | 755.00         |                |
| Exp. | 125.83  | 125.83 | 125.83 | 125.83  | 125.83 | 125.83 |                | <b>p-value</b> |
| X^2  | 123.84  | 19.21  | 121.87 | 1285.34 | 119.90 | 50.65  | <b>1720.81</b> | 4.074E-370     |
| CTG  | 462.00  | 1.00   | 34.00  | 2.00    | 0.00   | 0.00   | 499.00         |                |
| Exp. | 83.17   | 83.17  | 83.17  | 83.17   | 83.17  | 83.17  |                | <b>p-value</b> |
| X^2  | 1725.63 | 81.18  | 29.07  | 79.21   | 83.17  | 83.17  | <b>2081.42</b> | 2.681E-448     |
| CAG  | 0.00    | 0.00   | 0.00   | 176.00  | 50.00  | 327.00 | 553.00         |                |
| Exp. | 92.17   | 92.17  | 92.17  | 92.17   | 92.17  | 92.17  |                | <b>p-value</b> |
| X^2  | 92.17   | 92.17  | 92.17  | 76.25   | 19.29  | 598.34 | <b>970.38</b>  | 1.552E-207     |
| CGA  | 0.00    | 0.00   | 0.00   | 0.00    | 0.00   | 0.00   | 0.00           |                |

|                |         |        |        |        |         |        |         |                |
|----------------|---------|--------|--------|--------|---------|--------|---------|----------------|
| Exp.           | N/A     | N/A    | N/A    | N/A    | N/A     | N/A    |         | <b>p-value</b> |
| X <sup>2</sup> | N/A     | N/A    | N/A    | N/A    | N/A     | N/A    | N/A     | N/A            |
| CGG            | 1.00    | 0.00   | 110.00 | 1.00   | 1.00    | 0.00   | 113.00  |                |
| Exp.           | 18.83   | 18.83  | 18.83  | 18.83  | 18.83   | 18.83  |         | <b>p-value</b> |
| X <sup>2</sup> | 16.89   | 18.83  | 441.31 | 16.89  | 16.89   | 18.83  | 529.64  | 3.1915E-112    |
| CGC            | 0.00    | 1.00   | 1.00   | 96.00  | 0.00    | 16.00  | 114.00  |                |
| Exp.           | 19.00   | 19.00  | 19.00  | 19.00  | 19.00   | 19.00  |         | <b>p-value</b> |
| X <sup>2</sup> | 19.00   | 17.05  | 17.05  | 312.05 | 19.00   | 0.47   | 384.63  | 6.08251E-81    |
| CGT            | 86.00   | 15.00  | 93.00  | 89.00  | 5.00    | 17.00  | 305.00  |                |
| Exp.           | 50.83   | 50.83  | 50.83  | 50.83  | 50.83   | 50.83  |         | <b>p-value</b> |
| X <sup>2</sup> | 24.33   | 25.26  | 34.98  | 28.66  | 41.33   | 22.52  | 177.07  | 2.26494E-36    |
| TAA            | 6.00    | 82.00  | 6.00   | 376.00 | 16.00   | 26.00  | 512.00  |                |
| Exp.           | 85.33   | 85.33  | 85.33  | 85.33  | 85.33   | 85.33  |         | <b>p-value</b> |
| X <sup>2</sup> | 73.76   | 0.13   | 73.76  | 990.08 | 56.33   | 41.26  | 1235.31 | 6.5893E-265    |
| TAC            | 878.00  | 152.00 | 408.00 | 88.00  | 26.00   | 71.00  | 1623.00 |                |
| Exp.           | 270.50  | 270.50 | 270.50 | 270.50 | 270.50  | 270.50 |         | <b>p-value</b> |
| X <sup>2</sup> | 1364.35 | 51.91  | 69.89  | 123.13 | 221.00  | 147.14 | 1977.42 | 9.513E-426     |
| TCA            | 9.00    | 1.00   | 210.00 | 83.00  | 608.00  | 44.00  | 955.00  |                |
| Exp.           | 159.17  | 159.17 | 159.17 | 159.17 | 159.17  | 159.17 |         | <b>p-value</b> |
| X <sup>2</sup> | 141.68  | 157.17 | 16.23  | 36.45  | 1265.66 | 83.33  | 1700.52 | 1.019E-365     |
| TCC            | 28.00   | 84.00  | 175.00 | 29.00  | 259.00  | 16.00  | 591.00  |                |
| Exp.           | 98.50   | 98.50  | 98.50  | 98.50  | 98.50   | 98.50  |         | <b>p-value</b> |
| X <sup>2</sup> | 50.46   | 2.13   | 59.41  | 49.04  | 261.53  | 69.10  | 491.67  | 5.014E-104     |
| TCT            | 5.00    | 0.00   | 287.00 | 6.00   | 163.00  | 1.00   | 462.00  |                |
| Exp.           | 77.00   | 77.00  | 77.00  | 77.00  | 77.00   | 77.00  |         | <b>p-value</b> |
| X <sup>2</sup> | 67.32   | 77.00  | 572.73 | 65.47  | 96.05   | 75.01  | 953.58  | 6.7143E-204    |
| TCG            | 0.00    | 0.00   | 3.00   | 1.00   | 91.00   | 96.00  | 191.00  |                |
| Exp.           | 31.83   | 31.83  | 31.83  | 31.83  | 31.83   | 31.83  |         | <b>p-value</b> |
| X <sup>2</sup> | 31.83   | 31.83  | 26.12  | 29.86  | 109.97  | 129.34 | 358.96  | 2.06179E-75    |
| TAT            | 63.00   | 89.00  | 201.00 | 69.00  | 198.00  | 11.00  | 631.00  |                |
| Exp.           | 105.17  | 105.17 | 105.17 | 105.17 | 105.17  | 105.17 |         | <b>p-value</b> |
| X <sup>2</sup> | 16.91   | 2.49   | 87.33  | 12.44  | 81.95   | 84.32  | 285.42  | 1.36177E-59    |
| TTA            | 153.00  | 41.00  | 221.00 | 45.00  | 41.00   | 91.00  | 592.00  |                |
| Exp.           | 98.67   | 98.67  | 98.67  | 98.67  | 98.67   | 98.67  |         | <b>p-value</b> |
| X <sup>2</sup> | 29.92   | 33.70  | 151.68 | 29.19  | 33.70   | 0.60   | 278.79  | 3.6208E-58     |
| TTT            | 111.00  | 184.00 | 5.00   | 139.00 | 1.00    | 24.00  | 464.00  |                |
| Exp.           | 77.33   | 77.33  | 77.33  | 77.33  | 77.33   | 77.33  |         | <b>p-value</b> |
| X <sup>2</sup> | 14.66   | 147.13 | 67.66  | 49.17  | 75.35   | 36.78  | 390.74  | 2.93475E-82    |
| TTC            | 3.00    | 78.00  | 0.00   | 198.00 | 2.00    | 185.00 | 466.00  |                |
| Exp.           | 77.67   | 77.67  | 77.67  | 77.67  | 77.67   | 77.67  |         | <b>p-value</b> |
| X <sup>2</sup> | 71.78   | 0.00   | 77.67  | 186.44 | 73.72   | 148.33 | 557.94  | 2.4656E-118    |
| TTG            | 239.00  | 1.00   | 233.00 | 0.00   | 0.00    | 0.00   | 473.00  |                |
| Exp.           | 78.83   | 78.83  | 78.83  | 78.83  | 78.83   | 78.83  |         | <b>p-value</b> |
| X <sup>2</sup> | 325.41  | 76.85  | 301.49 | 78.83  | 78.83   | 78.83  | 940.25  | 5.1754E-201    |
| TAG            | 0.00    | 0.00   | 0.00   | 251.00 | 19.00   | 187.00 | 457.00  |                |
| Exp.           | 76.17   | 76.17  | 76.17  | 76.17  | 76.17   | 76.17  |         | <b>p-value</b> |
| X <sup>2</sup> | 76.17   | 76.17  | 76.17  | 401.31 | 42.91   | 161.28 | 834.00  | 5.1027E-178    |
| TGA            | 65.00   | 157.00 | 3.00   | 212.00 | 0.00    | 1.00   | 438.00  |                |
| Exp.           | 73.00   | 73.00  | 73.00  | 73.00  | 73.00   | 73.00  |         | <b>p-value</b> |
| X <sup>2</sup> | 0.88    | 96.66  | 67.12  | 264.67 | 73.00   | 71.01  | 573.34  | 1.1614E-121    |
| TGG            | 0.00    | 223.00 | 0.00   | 37.00  | 0.00    | 0.00   | 260.00  |                |
| Exp.           | 43.33   | 43.33  | 43.33  | 43.33  | 43.33   | 43.33  |         | <b>p-value</b> |
| X <sup>2</sup> | 43.33   | 744.93 | 43.33  | 0.93   | 43.33   | 43.33  | 919.18  | 1.8747E-196    |
| TGC            | 7.00    | 99.00  | 0.00   | 237.00 | 0.00    | 0.00   | 343.00  |                |
| Exp.           | 57.17   | 57.17  | 57.17  | 57.17  | 57.17   | 57.17  |         | <b>p-value</b> |
| X <sup>2</sup> | 44.02   | 30.61  | 57.17  | 565.71 | 57.17   | 57.17  | 811.85  | 3.1577E-173    |
| TGT            | 0.00    | 17.00  | 0.00   | 387.00 | 2.00    | 1.00   | 407.00  |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|                |        |        |        |         |        |        |                |                |
|----------------|--------|--------|--------|---------|--------|--------|----------------|----------------|
| Exp.           | 67.83  | 67.83  | 67.83  | 67.83   | 67.83  | 67.83  |                | <b>p-value</b> |
| X <sup>2</sup> | 67.83  | 38.09  | 67.83  | 1501.73 | 63.89  | 65.85  | <b>1805.23</b> | 2.040E-388     |
| GAA            | 90.00  | 5.00   | 3.00   | 4.00    | 49.00  | 58.00  | 209.00         |                |
| Exp.           | 34.83  | 34.83  | 34.83  | 34.83   | 34.83  | 34.83  |                | <b>p-value</b> |
| X <sup>2</sup> | 87.37  | 25.55  | 29.09  | 27.29   | 5.76   | 15.41  | <b>190.47</b>  | 3.09427E-39    |
| GAC            | 0.00   | 1.00   | 2.00   | 58.00   | 142.00 | 116.00 | 319.00         |                |
| Exp.           | 53.17  | 53.17  | 53.17  | 53.17   | 53.17  | 53.17  |                | <b>p-value</b> |
| X <sup>2</sup> | 53.17  | 51.19  | 49.24  | 0.44    | 148.43 | 74.26  | <b>376.72</b>  | 3.08355E-79    |
| GCA            | 142.00 | 108.00 | 37.00  | 94.00   | 90.00  | 104.00 | 575.00         |                |
| Exp.           | 95.83  | 95.83  | 95.83  | 95.83   | 95.83  | 95.83  |                | <b>p-value</b> |
| X <sup>2</sup> | 22.24  | 1.54   | 36.12  | 0.04    | 0.36   | 0.70   | <b>60.99</b>   | 7.58872E-12    |
| GCC            | 209.00 | 66.00  | 31.00  | 4.00    | 115.00 | 1.00   | 426.00         |                |
| Exp.           | 71.00  | 71.00  | 71.00  | 71.00   | 71.00  | 71.00  |                | <b>p-value</b> |
| X <sup>2</sup> | 268.23 | 0.35   | 22.54  | 63.23   | 27.27  | 69.01  | <b>450.62</b>  | 3.61072E-95    |
| GCT            | 0.00   | 0.00   | 32.00  | 7.00    | 78.00  | 100.00 | 217.00         |                |
| Exp.           | 36.17  | 36.17  | 36.17  | 36.17   | 36.17  | 36.17  |                | <b>p-value</b> |
| X <sup>2</sup> | 36.17  | 36.17  | 0.48   | 23.52   | 48.39  | 112.66 | <b>257.39</b>  | 1.42841E-53    |
| GCG            | 0.00   | 0.00   | 0.00   | 0.00    | 0.00   | 0.00   | 0.00           |                |
| Exp.           | N/A    | N/A    | N/A    | N/A     | N/A    | N/A    |                | <b>p-value</b> |
| X <sup>2</sup> | N/A    | N/A    | N/A    | N/A     | N/A    | N/A    | N/A            | N/A            |
| GAT            | 0.00   | 145.00 | 0.00   | 147.00  | 0.00   | 17.00  | 309.00         |                |
| Exp.           | 51.50  | 51.50  | 51.50  | 51.50   | 51.50  | 51.50  |                | <b>p-value</b> |
| X <sup>2</sup> | 51.50  | 169.75 | 51.50  | 177.09  | 51.50  | 23.11  | <b>524.46</b>  | 4.194E-111     |
| GTA            | 329.00 | 173.00 | 49.00  | 176.00  | 266.00 | 229.00 | 1222.00        |                |
| Exp.           | 203.67 | 203.67 | 203.67 | 203.67  | 203.67 | 203.67 |                | <b>p-value</b> |
| X <sup>2</sup> | 77.13  | 4.62   | 117.46 | 3.76    | 19.08  | 3.15   | <b>225.19</b>  | 1.14924E-46    |
| GTT            | 0.00   | 0.00   | 0.00   | 0.00    | 90.00  | 1.00   | 91.00          |                |
| Exp.           | 15.17  | 15.17  | 15.17  | 15.17   | 15.17  | 15.17  |                | <b>p-value</b> |
| X <sup>2</sup> | 15.17  | 15.17  | 15.17  | 15.17   | 369.23 | 13.23  | <b>443.13</b>  | 1.48831E-93    |
| GTC            | 85.00  | 73.00  | 1.00   | 12.00   | 0.00   | 37.00  | 208.00         |                |
| Exp.           | 34.67  | 34.67  | 34.67  | 34.67   | 34.67  | 34.67  |                | <b>p-value</b> |
| X <sup>2</sup> | 73.08  | 42.39  | 32.70  | 14.82   | 34.67  | 0.16   | <b>197.81</b>  | 8.36294E-41    |
| GTG            | 1.00   | 0.00   | 5.00   | 1.00    | 2.00   | 3.00   | 12.00          |                |
| Exp.           | 2.00   | 2.00   | 2.00   | 2.00    | 2.00   | 2.00   |                | <b>p-value</b> |
| X <sup>2</sup> | N/A    | N/A    | N/A    | N/A     | N/A    | N/A    | N/A            | N/A            |
| GAG            | 0.00   | 0.00   | 0.00   | 5.00    | 0.00   | 0.00   | 5.00           |                |
| Exp.           | 0.83   | 0.83   | 0.83   | 0.83    | 0.83   | 0.83   |                | <b>p-value</b> |
| X <sup>2</sup> | N/A    | N/A    | N/A    | N/A     | N/A    | N/A    | N/A            | N/A            |
| GGA            | 193.00 | 4.00   | 1.00   | 0.00    | 48.00  | 13.00  | 259.00         |                |
| Exp.           | 43.17  | 43.17  | 43.17  | 43.17   | 43.17  | 43.17  |                | <b>p-value</b> |
| X <sup>2</sup> | 520.08 | 35.54  | 41.19  | 43.17   | 0.54   | 21.08  | <b>661.59</b>  | 9.8681E-141    |
| GGG            | 14.00  | 29.00  | 0.00   | 1.00    | 0.00   | 197.00 | 241.00         |                |
| Exp.           | 40.17  | 40.17  | 40.17  | 40.17   | 40.17  | 40.17  |                | <b>p-value</b> |
| X <sup>2</sup> | 17.05  | 3.10   | 40.17  | 38.19   | 40.17  | 612.37 | <b>751.04</b>  | 4.5026E-160    |
| GGC            | 0.00   | 0.00   | 0.00   | 0.00    | 2.00   | 0.00   | 2.00           |                |
| Exp.           | 0.33   | 0.33   | 0.33   | 0.33    | 0.33   | 0.33   |                | <b>p-value</b> |
| X <sup>2</sup> | N/A    | N/A    | N/A    | N/A     | N/A    | N/A    | N/A            | N/A            |
| GGT            | 1.00   | 1.00   | 15.00  | 111.00  | 91.00  | 40.00  | 259.00         |                |
| Exp.           | 43.17  | 43.17  | 43.17  | 43.17   | 43.17  | 43.17  |                | <b>p-value</b> |
| X <sup>2</sup> | 41.19  | 41.19  | 18.38  | 106.60  | 53.00  | 0.23   | <b>260.59</b>  | 2.93217E-54    |

\*Kolmogorov test performed and generally consistent with the Chi Square test.

**Table B.** The p-values for each of the frames and each of the primers are provided. Any p-value < 7.8125E-4 (0.05/64) is considered statistically significant with an overall  $\alpha = 0.05$  for that frame. All p-values exceeding 7.8125E-4 are black reversed; the null hypothesis is not to be rejected. Overall, the frames are not distributed uniformly across the six patterns. All N/A correspond to no observations or too few observations for that frame.

**Laboratory A**

| Frame | p-value     |              |              |              |
|-------|-------------|--------------|--------------|--------------|
|       | A1          | B1           | C1           | D1           |
| AAA   | 1.10576E-69 | 1.97613E-16  | 7.3393E-121  | 1.12829E-10  |
| AAC   | 2.19786E-85 | 2.21986E-08* | 1.8946E-117  | 5.98581E-08* |
| ACA   | 4.94025E-77 | 3.49871E-18  | 6.0921E-227  | 7.98679E-37  |
| ACC   | 3.90081E-25 | 3.50665E-33  | 2.5634E-42   | 1.59481E-15* |
| ACT   | 7.5747E-192 | 1.32776E-67  | 6.09762E-67  | 5.86839E-09* |
| ACG   | 1.9299E-87  | 1.33474E-49  | 5.449E-153   | 4.71333E-08* |
| AAT   | 7.39751E-29 | 8.498E-35    | 2.2824E-194  | 7.33427E-07  |
| ATA   | 6.7716E-186 | 5.61172E-11  | 4.0275E-119  | 6.84257E-11  |
| ATT   | 4.21498E-67 | 3.29153E-24  | 4.0832E-286  | 8.60482E-08  |
| ATC   | 2.82496E-66 | 2.79912E-27  | 1.6738E-118  | 1.47321E-11  |
| ATG   | 9.4623E-202 | 2.93389E-62  | 2.8625E-201  | 9.20277E-21  |
| AAG   | 2.82071E-61 | 3.82785E-63  | 7.71945E-35  | 4.22105E-48  |
| AGA   | 8.7789E-151 | 1.11052E-35  | 4.8407E-206  | 8.05448E-85  |
| AGG   | 1.01943E-74 | 2.2631E-106  | 1.3894E-181  | 3.5831E-42   |
| AGC   | 4.6226E-100 | 1.16187E-27  | 6.3073E-203  | 4.8062E-43   |
| AGT   | 1.936E-173  | 2.79175E-96  | 9.97307E-67  | 1.52373E-35  |
| CAA   | 7.4958E-169 | 1.90125E-06  | 1.9824E-119  | 3.7819E-13   |
| CAC   | 2.1631E-109 | N/A          | 3.4639E-109  | 0.000455274  |
| CCA   | 2.599E-153  | 0.041940545  | 4.4458E-198  | 9.2056E-32   |
| CCC   | 3.4899E-265 | 3.985E-29    | 3.4124E-218  | 3.63988E-05* |
| CCT   | 2.9344E-134 | 0.000322535  | 6.49566E-54  | 2.05827E-09* |
| CCG   | 2.63361E-37 | 1.05168E-24* | 6.86065E-43  | 9.49811E-07* |
| CAT   | 5.091E-156  | 3.53958E-22  | 1.0019E-73   | 5.81374E-24  |
| CTA   | 1.2478E-107 | 1.04622E-19  | 4.00094E-87  | 1.23556E-06* |
| CTT   | 2.099E-142  | 1.33634E-56  | 1.57331E-90  | 6.95553E-18  |
| CTC   | 4.074E-370  | 1.89699E-09* | 1.82576E-65  | 1.1002E-24   |
| CTG   | 2.681E-448  | 6.54255E-38  | 1.6213E-223  | 2.34536E-24  |
| CAG   | 1.552E-207  | 1.70036E-47  | 2.8300E-370  | 1.22861E-93  |
| CGA   | N/A         | 1.44346E-42  | 1.1667E-162  | 8.58593E-20  |
| CGG   | 3.1915E-112 | 1.935E-08    | 2.81304E-48  | 6.04351E-32  |
| CGC   | 6.08251E-81 | N/A          | 4.44939E-71  | 1.14478E-09  |
| CGT   | 2.26494E-36 | 9.19214E-17* | 1.54625E-91  | 2.88296E-13  |
| TAA   | 6.5893E-265 | 1.68872E-31  | 6.0223E-238  | 6.16486E-08  |
| TAC   | 9.513E-426  | 8.22043E-88  | 4.2511E-148  | 2.2579E-14   |
| TCA   | 1.019E-365  | 8.54231E-19  | 2.9824E-217  | 1.08796E-22* |
| TCC   | 5.014E-104  | 1.65804E-36  | 4.67588E-79  | 2.91618E-25  |
| TCT   | 6.7143E-204 | 3.1875E-13   | 1.31091E-95  | 2.5745E-10   |
| TCG   | 2.06179E-75 | N/A          | 4.01349E-83  | 4.33358E-33  |
| TAT   | 1.36177E-59 | 2.09234E-36  | 1.9361E-130  | 1.23074E-25  |
| TTA   | 3.6208E-58  | 1.56213E-13  | 3.4131E-182  | 3.43921E-42  |
| TTT   | 2.93475E-82 | 1.3947E-81   | 6.6018E-230  | 1.0613E-29   |
| TTC   | 2.4656E-118 | 1.06789E-18  | 2.4707E-244  | 2.13543E-22  |
| TTG   | 5.1754E-201 | 8.63259E-42  | 2.6052E-177  | 3.64868E-65  |
| TAG   | 5.1027E-178 | 3.76452E-83  | 4.75852E-90  | 2.79722E-66  |
| TGA   | 1.1614E-121 | 4.58887E-79  | 1.1671E-140  | 1.28784E-06  |
| TGG   | 1.8747E-196 | 5.271E-137   | 6.65475E-85  | 4.48611E-84  |
| TGC   | 3.1577E-173 | 7.34499E-14  | 6.9029E-218  | 1.69061E-32  |
| TGT   | 2.040E-388  | 5.9162E-142  | <1.0048E-499 | 5.1127E-112  |
| GAA   | 3.09427E-39 | 6.61348E-35  | 1.1894E-100  | 3.04136E-24  |
| GAC   | 3.08355E-79 | 1.08369E-54  | 6.5042E-125  | 1.31739E-51  |
| GCA   | 7.58872E-12 | 1.7476E-13   | 2.90678E-57  | 2.21434E-13  |
| GCC   | 3.61072E-95 | N/A          | 4.5238E-85   | 1.06769E-15  |
| GCT   | 1.42841E-53 | 3.3107E-21   | 3.79508E-79  | 0.000609876  |
| GCG   | N/A         | 4.48654E-27* | 2.216E-119   | 6.1411E-12   |
| GAT   | 4.194E-111  | 1.31856E-59  | 1.5307E-107  | 1.78362E-30  |
| GTA   | 1.14924E-46 | 4.17879E-51  | 5.65438E-59  | 7.79196E-20  |
| GTT   | 1.48831E-93 | 1.05207E-35  | 5.02487E-90  | 4.17155E-38  |
| GTC   | 8.36294E-41 | 4.37348E-12  | 8.3542E-171  | 2.26187E-16  |
| GTG   | N/A         | 1.15875E-58  | 8.8089E-104  | 5.51488E-43  |
| GAG   | N/A         | 1.5569E-121  | 1.56937E-88  | 1.11912E-14  |
| GGA   | 9.8681E-141 | 1.49611E-24  | 8.47539E-40  | 0.033005783  |
| GGG   | 4.5026E-160 | 8.68089E-75  | 1.62366E-61  | 3.58748E-25  |
| GGC   | N/A         | 1.21759E-56  | 2.96643E-71  | 1.6383E-11   |
| GGT   | 2.93217E-54 | 3.99124E-22  | 1.06385E-85  | 1.56661E-30  |

\*Kolmogorov test performed and generally consistent with the Chi Square test.

### Laboratory B

| Frame | p-value      |              |             |              |
|-------|--------------|--------------|-------------|--------------|
|       | A1           | B1           | C1          | D1           |
| AAA   | 2.4015E-19   | 5.66898E-41  | 5.94755E-39 | 4.27164E-07  |
| AAC   | 1.47379E-40  | 4.66112E-21* | 1.20691E-67 | 1.41557E-08  |
| ACA   | 5.81954E-28  | 1.07704E-31  | 1.9407E-70  | 2.62816E-46  |
| ACC   | 4.5398E-08   | 4.15116E-55  | 4.08269E-18 | 2.22103E-32  |
| ACT   | 1.99246E-88  | 4.5485E-97   | 2.35893E-18 | 2.83706E-12  |
| ACG   | 7.72488E-40  | 6.43297E-41  | 8.07106E-43 | 2.37484E-07  |
| AAT   | 1.57917E-10  | 2.29438E-54  | 6.33869E-56 | 1.81591E-14  |
| ATA   | 3.08042E-42  | 2.18098E-11  | 1.91897E-21 | 1.10885E-19  |
| ATT   | 1.42826E-49  | 9.18013E-20  | 2.39761E-61 | 4.76423E-09  |
| ATC   | 4.51011E-41  | 3.05379E-35  | 8.81307E-56 | 1.23928E-12  |
| ATG   | 1.98679E-77  | 4.09257E-32  | 2.90208E-57 | 7.84509E-24  |
| AAG   | 2.33415E-25  | 9.3961E-61   | 0.01188574  | 1.7849E-48   |
| AGA   | 8.42606E-64  | 2.09949E-45  | 3.99923E-82 | 5.5188E-100  |
| AGG   | 9.59874E-25  | 3.2893E-137  | 5.23552E-41 | 1.18655E-41  |
| AGC   | 2.92461E-30  | 5.01139E-34  | 1.75198E-42 | 1.23857E-47  |
| AGT   | 4.72293E-47  | 1.5905E-149  | 2.26133E-15 | 6.06819E-39  |
| CAA   | 9.77873E-50  | 3.16667E-16  | 8.02374E-32 | 8.94217E-18  |
| CAC   | 1.08711E-28  | N/A          | 3.10063E-27 | 3.51349E-12  |
| CCA   | 4.58102E-35  | 3.27143E-08  | 1.5628E-46  | 2.26276E-31  |
| CCC   | 4.4424E-111  | 1.02667E-16  | 1.09157E-62 | 1.24518E-07* |
| CCT   | 1.27408E-58  | 2.31622E-08  | 6.53436E-21 | 1.35044E-30  |
| CCG   | 1.72067E-22  | 8.05057E-30  | 1.68456E-15 | 0.000978458* |
| CAT   | 2.77056E-44  | 2.27913E-37  | 1.0623E-24  | 3.06673E-34  |
| CTA   | 2.99906E-42  | 1.40316E-08  | 5.59292E-13 | 7.5511E-12   |
| CTT   | 4.10768E-42  | 3.46188E-91  | 2.27655E-11 | 2.0618E-44   |
| CTC   | 2.2497E-117  | N/A          | 5.28302E-13 | 4.73764E-40  |
| CTG   | 2.103E-150   | 4.04892E-43  | 8.80036E-65 | 1.20559E-33  |
| CAG   | 9.45356E-65  | 1.19815E-55  | 2.592E-101  | 1.21807E-87  |
| CGA   | N/A          | 2.33428E-27  | 2.95234E-42 | 2.87206E-26  |
| CGG   | 4.35835E-35  | 9.75139E-07  | 5.45893E-20 | 2.21247E-35  |
| CGC   | 2.55874E-30  | N/A          | 9.26587E-19 | 1.31116E-14  |
| CGT   | 3.97319E-25  | 1.87527E-23  | 1.21979E-16 | 4.45073E-07  |
| TAA   | 2.06078E-64  | 3.90672E-39  | 5.5774E-55  | 1.91901E-05  |
| TAC   | 9.5631E-174  | 2.29264E-85  | 2.48687E-48 | 1.40833E-24  |
| TCA   | 1.2769E-273  | 7.14309E-21  | 4.42298E-39 | 7.07389E-25  |
| TCC   | 2.21154E-30  | 5.90659E-63  | 1.18035E-13 | 2.94546E-14  |
| TCT   | 2.05411E-48  | 1.15443E-20  | 1.30145E-33 | 4.08338E-36  |
| TCG   | 5.76146E-60  | N/A          | 7.24764E-29 | 2.26474E-45  |
| TAT   | 2.15884E-38  | 7.51926E-78  | 1.18758E-51 | 1.26286E-21  |
| TTA   | 1.72504E-47  | 1.89386E-33  | 2.09194E-57 | 8.29455E-51  |
| TTT   | 1.07355E-32  | 3.1084E-163  | 7.04152E-45 | 5.66438E-36  |
| TTC   | 5.81169E-38  | 4.22246E-45  | 5.3141E-65  | 1.05579E-18  |
| TTG   | 5.90393E-84  | 1.43254E-68  | 5.30689E-65 | 8.99449E-72  |
| TAG   | 1.51082E-57  | 1.8319E-98   | 4.408E-43   | 1.78966E-69  |
| TGA   | 5.65214E-44  | 1.1231E-130  | 8.36594E-69 | 6.69995E-09  |
| TGG   | 1.05602E-42  | 3.9485E-158  | 2.22348E-29 | 5.38629E-98  |
| TGC   | 7.56442E-60  | 5.28634E-27  | 2.26696E-53 | 1.88197E-54  |
| TGT   | 1.164E-140   | 5.8607E-244  | 2.5967E-187 | 7.3112E-157  |
| GAA   | 2.40547E-20  | 1.43418E-51  | 3.45816E-26 | 4.96895E-28  |
| GAC   | 2.36184E-29  | 8.7428E-50   | 9.84828E-33 | 1.29791E-70  |
| GCA   | 0.020564676  | 7.24901E-15  | 5.24932E-13 | 5.5385E-15   |
| GCC   | 3.34127E-56  | N/A          | 1.11361E-23 | 3.59097E-23  |
| GCT   | 1.08487E-26  | 3.77338E-35  | 3.81025E-25 | 7.69814E-24  |
| GCG   | N/A          | 4.20543E-24  | 8.17806E-31 | 1.4237E-10   |
| GAT   | 2.12048E-44  | 1.98722E-58  | 1.94885E-34 | 1.24928E-65  |
| GTA   | 4.86965E-19  | 3.02542E-81  | 2.0574E-18  | 4.31008E-07  |
| GTT   | 4.09733E-22* | 5.73242E-46  | 1.77089E-19 | 9.06237E-88  |
| GTC   | 2.99389E-06  | 1.83455E-23  | 1.93039E-40 | 7.57482E-51  |
| GTG   | N/A          | 5.33664E-93  | 3.01566E-34 | 8.51628E-76  |
| GAG   | N/A          | 2.61428E-91  | 1.6611E-16  | 6.39698E-20  |
| GGA   | 7.96391E-45  | 2.13174E-26  | 7.5511E-12  | 1.76546E-08  |
| GGG   | 4.40048E-44  | 2.74661E-65  | 2.4203E-15* | 1.05373E-44  |
| GGC   | N/A          | 2.03016E-96  | 6.27076E-24 | 2.85272E-14  |
| GGT   | 5.36344E-22  | 4.60842E-40  | 4.17504E-31 | 5.84254E-89  |

\*Kolmogorov test performed and generally consistent with the Chi Square test.

### Appendix III — Comparison of Frames for Laboratory A to Laboratory B for Each Primer

The following table is an example of the processed data using the bioinformatics tools to statistically evaluate each of the 64 frames using a 2 x 6 Chi Square analysis comparing the results from Laboratory A to Laboratory B. For those frames with expected values less than that required in more than 25% of the cells, a Kolmogorov-Smirnov test was performed. These tables are not presented here; the p-values calculated with the Kolmogorov-Smirnov test statistic were consistent with the Chi Square.

These results demonstrate that the data obtained from the two laboratories are equivalent. Each of the four primers, A1, B1, C1, and D1, were evaluated to characterize the pattern distributions for each frame.

**Table A.** 2 x 6 Chi-square analysis comparing the results from Laboratory A to Laboratory B for Primer A1.

|                |               |               |               |               |               |               |                |                |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|
|                | <b>A</b>      | <b>B</b>      | <b>C</b>      | <b>D</b>      | <b>E</b>      | <b>F</b>      | <b>TOTAL</b>   |                |
| AAA            | 199.00        | 87.00         | 264.00        | 1.00          | 159.00        | 85.00         | 795.00         |                |
| AAA            | 71.00         | 31.00         | 76.00         | 0.00          | 58.00         | 29.00         | 265.00         |                |
| <b>TOTAL</b>   | <b>270.00</b> | <b>118.00</b> | <b>340.00</b> | <b>1.00</b>   | <b>217.00</b> | <b>114.00</b> | <b>1060.00</b> |                |
| Exp.           | 202.50        | 88.50         | 255.00        | 0.75          | 162.75        | 85.50         |                | <b>p-value</b> |
| Exp.           | 67.50         | 29.50         | 85.00         | 0.25          | 54.25         | 28.50         |                | 0.805545483    |
| X <sup>2</sup> | 0.06          | 0.03          | 0.32          | 0.08          | 0.09          | 0.00          | <b>0.58</b>    |                |
| X <sup>2</sup> | 0.18          | 0.08          | 0.95          | 0.25          | 0.26          | 0.01          | <b>1.73</b>    |                |
| AAC            | 296.00        | 271.00        | 353.00        | 26.00         | 217.00        | 75.00         | 1238.00        |                |
| AAC            | 92.00         | 74.00         | 152.00        | 15.00         | 76.00         | 11.00         | 420.00         |                |
| <b>TOTAL</b>   | <b>388.00</b> | <b>345.00</b> | <b>505.00</b> | <b>41.00</b>  | <b>293.00</b> | <b>86.00</b>  | <b>1658.00</b> |                |
| Exp.           | 289.71        | 257.61        | 377.07        | 30.61         | 218.78        | 64.21         |                | <b>p-value</b> |
| Exp.           | 98.29         | 87.39         | 127.93        | 10.39         | 74.22         | 21.79         |                | 0.001683344    |
| X <sup>2</sup> | 0.14          | 0.70          | 1.54          | 0.70          | 0.01          | 1.81          | <b>4.89</b>    |                |
| X <sup>2</sup> | 0.40          | 2.05          | 4.53          | 2.05          | 0.04          | 5.34          | <b>14.42</b>   |                |
| ACA            | 174.00        | 330.00        | 414.00        | 380.00        | 56.00         | 204.00        | 1558.00        |                |
| ACA            | 87.00         | 134.00        | 114.00        | 131.00        | 8.00          | 55.00         | 529.00         |                |
| <b>TOTAL</b>   | <b>261.00</b> | <b>464.00</b> | <b>528.00</b> | <b>511.00</b> | <b>64.00</b>  | <b>259.00</b> | <b>2087.00</b> |                |
| Exp.           | 194.84        | 346.39        | 394.17        | 381.47        | 47.78         | 193.35        |                | <b>p-value</b> |
| Exp.           | 66.16         | 117.61        | 133.83        | 129.53        | 16.22         | 65.65         |                | 0.000246578    |
| X <sup>2</sup> | 2.23          | 0.78          | 1.00          | 0.01          | 1.42          | 0.59          | <b>6.01</b>    |                |
| X <sup>2</sup> | 6.57          | 2.28          | 2.94          | 0.02          | 4.17          | 1.73          | <b>17.70</b>   |                |
| ACC            | 269.00        | 503.00        | 386.00        | 318.00        | 288.00        | 268.00        | 2032.00        |                |
| ACC            | 126.00        | 142.00        | 153.00        | 81.00         | 128.00        | 77.00         | 707.00         |                |
| <b>TOTAL</b>   | <b>395.00</b> | <b>645.00</b> | <b>539.00</b> | <b>399.00</b> | <b>416.00</b> | <b>345.00</b> | <b>2739.00</b> |                |
| Exp.           | 293.04        | 478.51        | 399.87        | 296.01        | 308.62        | 255.95        |                | <b>p-value</b> |
| Exp.           | 101.96        | 166.49        | 139.13        | 102.99        | 107.38        | 89.05         |                | 3.28633E-05    |
| X <sup>2</sup> | 1.97          | 1.25          | 0.48          | 1.63          | 1.38          | 0.57          | <b>7.29</b>    |                |
| X <sup>2</sup> | 5.67          | 3.60          | 1.38          | 4.70          | 3.96          | 1.63          | <b>20.94</b>   |                |
| ACT            | 445.00        | 58.00         | 98.00         | 104.00        | 112.00        | 2.00          | 819.00         |                |
| ACT            | 172.00        | 26.00         | 8.00          | 37.00         | 42.00         | 0.00          | 285.00         |                |
| <b>TOTAL</b>   | <b>617.00</b> | <b>84.00</b>  | <b>106.00</b> | <b>141.00</b> | <b>154.00</b> | <b>2.00</b>   | <b>1104.00</b> |                |
| Exp.           | 457.72        | 62.32         | 78.64         | 104.60        | 114.24        | 1.48          |                | <b>p-value</b> |
| Exp.           | 159.28        | 21.68         | 27.36         | 36.40         | 39.76         | 0.52          |                | 0.000552211    |
| X <sup>2</sup> | 0.35          | 0.30          | 4.77          | 0.00          | 0.04          | 0.18          | <b>5.65</b>    |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|              |               |               |               |               |               |               |                |                |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|
| X^2          | 1.02          | 0.86          | 13.70         | 0.01          | 0.13          | 0.52          | <b>16.23</b>   |                |
| ACG          | 1.00          | 98.00         | 0.00          | 11.00         | 0.00          | 1.00          | 111.00         |                |
| ACG          | 1.00          | 40.00         | 0.00          | 0.00          | 0.00          | 0.00          | 41.00          |                |
| <b>TOTAL</b> | <b>2.00</b>   | <b>138.00</b> | <b>0.00</b>   | <b>11.00</b>  | <b>0.00</b>   | <b>1.00</b>   | <b>152.00</b>  |                |
| Exp.         | 1.46          | 100.78        | 0.25          | 8.03          | 0.25          | 0.73          |                | <b>p-value</b> |
| Exp.         | 0.54          | 37.22         | 0.25          | 2.97          | 0.25          | 0.27          |                | 0.38563627*    |
| X^2          | 0.15          | 0.08          | 0.25          | 1.10          | 0.25          | 0.10          | <b>1.92</b>    |                |
| X^2          | 0.39          | 0.21          | 0.25          | 2.97          | 0.25          | 0.27          | <b>4.34</b>    |                |
| AAT          | 13.00         | 120.00        | 81.00         | 66.00         | 162.00        | 110.00        | 552.00         |                |
| AAT          | 5.00          | 42.00         | 46.00         | 31.00         | 54.00         | 16.00         | 194.00         |                |
| <b>TOTAL</b> | <b>18.00</b>  | <b>162.00</b> | <b>127.00</b> | <b>97.00</b>  | <b>216.00</b> | <b>126.00</b> | <b>746.00</b>  |                |
| Exp.         | 13.32         | 119.87        | 93.97         | 71.77         | 159.83        | 93.23         |                | <b>p-value</b> |
| Exp.         | 4.68          | 42.13         | 33.03         | 25.23         | 56.17         | 32.77         |                | 0.001045754    |
| X^2          | 0.01          | 0.00          | 1.79          | 0.46          | 0.03          | 3.02          | <b>5.31</b>    |                |
| X^2          | 0.02          | 0.00          | 5.10          | 1.32          | 0.08          | 8.58          | <b>15.10</b>   |                |
| ATA          | 185.00        | 94.00         | 148.00        | 5.00          | 445.00        | 18.00         | 895.00         |                |
| ATA          | 68.00         | 34.00         | 49.00         | 5.00          | 139.00        | 27.00         | 322.00         |                |
| <b>TOTAL</b> | <b>253.00</b> | <b>128.00</b> | <b>197.00</b> | <b>10.00</b>  | <b>584.00</b> | <b>45.00</b>  | <b>1217.00</b> |                |
| Exp.         | 186.06        | 94.13         | 144.88        | 7.35          | 429.48        | 33.09         |                | <b>p-value</b> |
| Exp.         | 66.94         | 33.87         | 52.12         | 2.65          | 154.52        | 11.91         |                | 8.30891E-06    |
| X^2          | 0.01          | 0.00          | 0.07          | 0.75          | 0.56          | 6.88          | <b>8.27</b>    |                |
| X^2          | 0.02          | 0.00          | 0.19          | 2.09          | 1.56          | 19.13         | <b>22.99</b>   |                |
| ATT          | 249.00        | 199.00        | 26.00         | 70.00         | 238.00        | 82.00         | 864.00         |                |
| ATT          | 23.00         | 99.00         | 6.00          | 32.00         | 7.00          | 113.00        | 280.00         |                |
| <b>TOTAL</b> | <b>272.00</b> | <b>298.00</b> | <b>32.00</b>  | <b>102.00</b> | <b>245.00</b> | <b>195.00</b> | <b>1144.00</b> |                |
| Exp.         | 205.43        | 225.06        | 24.17         | 77.03         | 185.03        | 147.27        |                | <b>p-value</b> |
| Exp.         | 66.57         | 72.94         | 7.83          | 24.97         | 59.97         | 47.73         |                | 1.97104E-48    |
| X^2          | 9.24          | 3.02          | 0.14          | 0.64          | 15.16         | 28.93         | <b>57.13</b>   |                |
| X^2          | 28.52         | 9.31          | 0.43          | 1.98          | 46.78         | 89.27         | <b>176.29</b>  |                |
| ATC          | 93.00         | 222.00        | 53.00         | 231.00        | 13.00         | 170.00        | 782.00         |                |
| ATC          | 38.00         | 74.00         | 7.00          | 118.00        | 0.00          | 56.00         | 293.00         |                |
| <b>TOTAL</b> | <b>131.00</b> | <b>296.00</b> | <b>60.00</b>  | <b>349.00</b> | <b>13.00</b>  | <b>226.00</b> | <b>1075.00</b> |                |
| Exp.         | 95.29         | 215.32        | 43.65         | 253.88        | 9.46          | 164.40        |                | <b>p-value</b> |
| Exp.         | 35.71         | 80.68         | 16.35         | 95.12         | 3.54          | 61.60         |                | 0.000665719    |
| X^2          | 0.06          | 0.21          | 2.00          | 2.06          | 1.33          | 0.19          | <b>5.85</b>    |                |
| X^2          | 0.15          | 0.55          | 5.35          | 5.50          | 3.54          | 0.51          | <b>15.60</b>   |                |
| ATG          | 193.00        | 0.00          | 267.00        | 0.00          | 3.00          | 0.00          | 463.00         |                |
| ATG          | 86.00         | 1.00          | 102.00        | 0.00          | 2.00          | 0.00          | 191.00         |                |
| <b>TOTAL</b> | <b>279.00</b> | <b>1.00</b>   | <b>369.00</b> | <b>0.00</b>   | <b>5.00</b>   | <b>0.00</b>   | <b>654.00</b>  |                |
| Exp.         | 197.52        | 0.71          | 261.23        | 0.25          | 3.54          | 0.25          |                | <b>p-value</b> |
| Exp.         | 81.48         | 0.29          | 107.77        | 0.25          | 1.46          | 0.25          |                | 0.624053924*   |
| X^2          | 0.10          | 0.71          | 0.13          | 0.25          | 0.08          | 0.25          | <b>1.52</b>    |                |
| X^2          | 0.25          | 1.72          | 0.31          | 0.25          | 0.20          | 0.25          | <b>2.97</b>    |                |
| AAG          | 73.00         | 176.00        | 26.00         | 219.00        | 54.00         | 61.00         | 609.00         |                |
| AAG          | 25.00         | 63.00         | 6.00          | 75.00         | 44.00         | 2.00          | 215.00         |                |
| <b>TOTAL</b> | <b>98.00</b>  | <b>239.00</b> | <b>32.00</b>  | <b>294.00</b> | <b>98.00</b>  | <b>63.00</b>  | <b>824.00</b>  |                |
| Exp.         | 72.43         | 176.64        | 23.65         | 217.29        | 72.43         | 46.56         |                | <b>p-value</b> |
| Exp.         | 25.57         | 62.36         | 8.35          | 76.71         | 25.57         | 16.44         |                | 9.05752E-07    |
| X^2          | 0.00          | 0.00          | 0.23          | 0.01          | 4.69          | 4.48          | <b>9.42</b>    |                |
| X^2          | 0.01          | 0.01          | 0.66          | 0.04          | 13.28         | 12.68         | <b>26.68</b>   |                |
| AGA          | 1.00          | 0.00          | 143.00        | 0.00          | 0.00          | 0.00          | 144.00         |                |
| AGA          | 0.00          | 0.00          | 61.00         | 0.00          | 0.00          | 0.00          | 61.00          |                |
| <b>TOTAL</b> | <b>1.00</b>   | <b>0.00</b>   | <b>204.00</b> | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>205.00</b>  |                |
| Exp.         | 0.70          | 0.25          | 143.30        | 0.25          | 0.25          | 0.25          |                | <b>p-value</b> |
| Exp.         | 0.30          | 0.25          | 60.70         | 0.25          | 0.25          | 0.25          |                | 0.994592443*   |
| X^2          | 0.13          | 0.25          | 0.00          | 0.25          | 0.25          | 0.25          | <b>1.13</b>    |                |
| X^2          | 0.30          | 0.25          | 0.00          | 0.25          | 0.25          | 0.25          | <b>1.30</b>    |                |
| AGG          | 0.00          | 0.00          | 46.00         | 0.00          | 101.00        | 0.00          | 147.00         |                |
| AGG          | 1.00          | 1.00          | 24.00         | 0.00          | 37.00         | 0.00          | 63.00          |                |
| <b>TOTAL</b> | <b>1.00</b>   | <b>1.00</b>   | <b>70.00</b>  | <b>0.00</b>   | <b>138.00</b> | <b>0.00</b>   | <b>210.00</b>  |                |
| Exp.         | 0.70          | 0.70          | 49.00         | 0.25          | 96.60         | 0.25          |                | <b>p-value</b> |
| Exp.         | 0.30          | 0.30          | 21.00         | 0.25          | 41.40         | 0.25          |                | 0.311414387*   |
| X^2          | 0.70          | 0.70          | 0.18          | 0.25          | 0.20          | 0.25          | <b>2.28</b>    |                |
| X^2          | 1.63          | 1.63          | 0.43          | 0.25          | 0.47          | 0.25          | <b>4.66</b>    |                |
| AGC          | 96.00         | 32.00         | 281.00        | 17.00         | 240.00        | 92.00         | 758.00         |                |
| AGC          | 47.00         | 3.00          | 91.00         | 6.00          | 71.00         | 32.00         | 250.00         |                |
| <b>TOTAL</b> | <b>143.00</b> | <b>35.00</b>  | <b>372.00</b> | <b>23.00</b>  | <b>311.00</b> | <b>124.00</b> | <b>1008.00</b> |                |



|              |               |               |               |               |               |               |                |  |                |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|--|----------------|
| Exp.         | 107.53        | 26.32         | 279.74        | 17.30         | 233.87        | 93.25         |                |  | <b>p-value</b> |
| Exp.         | 35.47         | 8.68          | 92.26         | 5.70          | 77.13         | 30.75         |                |  | 0.057880438    |
| X^2          | 1.24          | 1.23          | 0.01          | 0.01          | 0.16          | 0.02          | <b>2.65</b>    |  |                |
| X^2          | 3.75          | 3.72          | 0.02          | 0.02          | 0.49          | 0.05          | <b>8.04</b>    |  |                |
| AGT          | 75.00         | 0.00          | 186.00        | 1.00          | 305.00        | 9.00          | 576.00         |  |                |
| AGT          | 56.00         | 4.00          | 28.00         | 1.00          | 105.00        | 15.00         | 209.00         |  |                |
| <b>TOTAL</b> | <b>131.00</b> | <b>4.00</b>   | <b>214.00</b> | <b>2.00</b>   | <b>410.00</b> | <b>24.00</b>  | <b>785.00</b>  |  |                |
| Exp.         | 96.12         | 2.94          | 157.02        | 1.47          | 300.84        | 17.61         |                |  | <b>p-value</b> |
| Exp.         | 34.88         | 1.06          | 56.98         | 0.53          | 109.16        | 6.39          |                |  | 1.05449E-12*   |
| X^2          | 4.64          | 2.94          | 5.35          | 0.15          | 0.06          | 4.21          | <b>17.34</b>   |  |                |
| X^2          | 12.79         | 8.09          | 14.74         | 0.41          | 0.16          | 11.60         | <b>47.79</b>   |  |                |
| CAA          | 75.00         | 261.00        | 216.00        | 427.00        | 617.00        | 81.00         | 1677.00        |  |                |
| CAA          | 29.00         | 110.00        | 72.00         | 152.00        | 202.00        | 31.00         | 596.00         |  |                |
| <b>TOTAL</b> | <b>104.00</b> | <b>371.00</b> | <b>288.00</b> | <b>579.00</b> | <b>819.00</b> | <b>112.00</b> | <b>2273.00</b> |  |                |
| Exp.         | 76.73         | 273.72        | 212.48        | 427.18        | 604.25        | 82.63         |                |  | <b>p-value</b> |
| Exp.         | 27.27         | 97.28         | 75.52         | 151.82        | 214.75        | 29.37         |                |  | 0.582330595    |
| X^2          | 0.04          | 0.59          | 0.06          | 0.00          | 0.27          | 0.03          | <b>0.99</b>    |  |                |
| X^2          | 0.11          | 1.66          | 0.16          | 0.00          | 0.76          | 0.09          | <b>2.78</b>    |  |                |
| CAC          | 126.00        | 481.00        | 30.00         | 270.00        | 213.00        | 222.00        | 1342.00        |  |                |
| CAC          | 50.00         | 143.00        | 5.00          | 92.00         | 75.00         | 80.00         | 445.00         |  |                |
| <b>TOTAL</b> | <b>176.00</b> | <b>624.00</b> | <b>35.00</b>  | <b>362.00</b> | <b>288.00</b> | <b>302.00</b> | <b>1787.00</b> |  |                |
| Exp.         | 132.17        | 468.61        | 26.28         | 271.85        | 216.28        | 226.80        |                |  | <b>p-value</b> |
| Exp.         | 43.83         | 155.39        | 8.72          | 90.15         | 71.72         | 75.20         |                |  | 0.387273169    |
| X^2          | 0.29          | 0.33          | 0.53          | 0.01          | 0.05          | 0.10          | <b>1.30</b>    |  |                |
| X^2          | 0.87          | 0.99          | 1.58          | 0.04          | 0.15          | 0.31          | <b>3.94</b>    |  |                |
| CCA          | 145.00        | 120.00        | 313.00        | 422.00        | 276.00        | 716.00        | 1992.00        |  |                |
| CCA          | 81.00         | 47.00         | 90.00         | 152.00        | 93.00         | 221.00        | 684.00         |  |                |
| <b>TOTAL</b> | <b>226.00</b> | <b>167.00</b> | <b>403.00</b> | <b>574.00</b> | <b>369.00</b> | <b>937.00</b> | <b>2676.00</b> |  |                |
| Exp.         | 168.23        | 124.31        | 299.99        | 427.28        | 274.68        | 697.50        |                |  | <b>p-value</b> |
| Exp.         | 57.77         | 42.69         | 103.01        | 146.72        | 94.32         | 239.50        |                |  | 0.003573017    |
| X^2          | 3.21          | 0.15          | 0.56          | 0.07          | 0.01          | 0.49          | <b>4.48</b>    |  |                |
| X^2          | 9.34          | 0.44          | 1.64          | 0.19          | 0.02          | 1.43          | <b>13.06</b>   |  |                |
| CCC          | 18.00         | 366.00        | 7.00          | 520.00        | 8.00          | 378.00        | 1297.00        |  |                |
| CCC          | 6.00          | 150.00        | 0.00          | 211.00        | 1.00          | 93.00         | 461.00         |  |                |
| <b>TOTAL</b> | <b>24.00</b>  | <b>516.00</b> | <b>7.00</b>   | <b>731.00</b> | <b>9.00</b>   | <b>471.00</b> | <b>1758.00</b> |  |                |
| Exp.         | 17.71         | 380.69        | 5.16          | 539.31        | 6.64          | 347.49        |                |  | <b>p-value</b> |
| Exp.         | 6.29          | 135.31        | 1.84          | 191.69        | 2.36          | 123.51        |                |  | 0.002298114    |
| X^2          | 0.00          | 0.57          | 0.65          | 0.69          | 0.28          | 2.68          | <b>4.87</b>    |  |                |
| X^2          | 0.01          | 1.59          | 1.84          | 1.95          | 0.78          | 7.54          | <b>13.71</b>   |  |                |
| CCT          | 98.00         | 29.00         | 312.00        | 37.00         | 275.00        | 36.00         | 787.00         |  |                |
| CCT          | 41.00         | 20.00         | 127.00        | 7.00          | 108.00        | 3.00          | 306.00         |  |                |
| <b>TOTAL</b> | <b>139.00</b> | <b>49.00</b>  | <b>439.00</b> | <b>44.00</b>  | <b>383.00</b> | <b>39.00</b>  | <b>1093.00</b> |  |                |
| Exp.         | 100.09        | 35.28         | 316.10        | 31.68         | 275.77        | 28.08         |                |  | <b>p-value</b> |
| Exp.         | 38.91         | 13.72         | 122.90        | 12.32         | 107.23        | 10.92         |                |  | 0.008383757    |
| X^2          | 0.04          | 1.12          | 0.05          | 0.89          | 0.00          | 2.23          | <b>4.34</b>    |  |                |
| X^2          | 0.11          | 2.88          | 0.14          | 2.30          | 0.01          | 5.74          | <b>11.17</b>   |  |                |
| CCG          | 93.00         | 12.00         | 25.00         | 4.00          | 78.00         | 19.00         | 231.00         |  |                |
| CCG          | 40.00         | 1.00          | 6.00          | 1.00          | 21.00         | 0.00          | 69.00          |  |                |
| <b>TOTAL</b> | <b>133.00</b> | <b>13.00</b>  | <b>31.00</b>  | <b>5.00</b>   | <b>99.00</b>  | <b>19.00</b>  | <b>300.00</b>  |  |                |
| Exp.         | 102.41        | 10.01         | 23.87         | 3.85          | 76.23         | 14.63         |                |  | <b>p-value</b> |
| Exp.         | 30.59         | 2.99          | 7.13          | 1.15          | 22.77         | 4.37          |                |  | 0.040836662    |
| X^2          | 0.86          | 0.40          | 0.05          | 0.01          | 0.04          | 1.31          | <b>2.67</b>    |  |                |
| X^2          | 2.89          | 1.32          | 0.18          | 0.02          | 0.14          | 4.37          | <b>8.93</b>    |  |                |
| CAT          | 43.00         | 368.00        | 217.00        | 245.00        | 555.00        | 71.00         | 1499.00        |  |                |
| CAT          | 48.00         | 144.00        | 72.00         | 63.00         | 184.00        | 23.00         | 534.00         |  |                |
| <b>TOTAL</b> | <b>91.00</b>  | <b>512.00</b> | <b>289.00</b> | <b>308.00</b> | <b>739.00</b> | <b>94.00</b>  | <b>2033.00</b> |  |                |
| Exp.         | 67.10         | 377.52        | 213.09        | 227.10        | 544.89        | 69.31         |                |  | <b>p-value</b> |
| Exp.         | 23.90         | 134.48        | 75.91         | 80.90         | 194.11        | 24.69         |                |  | 1.25314E-07    |
| X^2          | 8.65          | 0.24          | 0.07          | 1.41          | 0.19          | 0.04          | <b>10.61</b>   |  |                |
| X^2          | 24.29         | 0.67          | 0.20          | 3.96          | 0.53          | 0.12          | <b>29.77</b>   |  |                |
| CTA          | 212.00        | 194.00        | 4.00          | 55.00         | 3.00          | 47.00         | 515.00         |  |                |
| CTA          | 82.00         | 52.00         | 1.00          | 27.00         | 1.00          | 2.00          | 165.00         |  |                |
| <b>TOTAL</b> | <b>294.00</b> | <b>246.00</b> | <b>5.00</b>   | <b>82.00</b>  | <b>4.00</b>   | <b>49.00</b>  | <b>680.00</b>  |  |                |
| Exp.         | 222.66        | 186.31        | 3.79          | 62.10         | 3.03          | 37.11         |                |  | <b>p-value</b> |
| Exp.         | 71.34         | 59.69         | 1.21          | 19.90         | 0.97          | 11.89         |                |  | 0.003385636*   |
| X^2          | 0.51          | 0.32          | 0.01          | 0.81          | 0.00          | 2.64          | <b>4.29</b>    |  |                |
| X^2          | 1.59          | 0.99          | 0.04          | 2.54          | 0.00          | 8.23          | <b>13.38</b>   |  |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|              |                |               |               |               |               |               |                |                |
|--------------|----------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|
| CTT          | 282.00         | 11.00         | 19.00         | 0.00          | 145.00        | 76.00         | 533.00         |                |
| CTT          | 98.00          | 8.00          | 4.00          | 3.00          | 58.00         | 43.00         | 214.00         |                |
| <b>TOTAL</b> | <b>380.00</b>  | <b>19.00</b>  | <b>23.00</b>  | <b>3.00</b>   | <b>203.00</b> | <b>119.00</b> | <b>747.00</b>  |                |
| Exp.         | 271.14         | 13.56         | 16.41         | 2.14          | 144.84        | 84.91         |                | <b>p-value</b> |
| Exp.         | 108.86         | 5.44          | 6.59          | 0.86          | 58.16         | 34.09         |                | 0.008916897*   |
| X^2          | 0.44           | 0.48          | 0.41          | 2.14          | 0.00          | 0.93          | <b>4.40</b>    |                |
| X^2          | 1.08           | 1.20          | 1.02          | 5.33          | 0.00          | 2.33          | <b>10.96</b>   |                |
| CTC          | 1.00           | 175.00        | 2.00          | 528.00        | 3.00          | 46.00         | 755.00         |                |
| CTC          | 0.00           | 62.00         | 4.00          | 174.00        | 0.00          | 15.00         | 255.00         |                |
| <b>TOTAL</b> | <b>1.00</b>    | <b>237.00</b> | <b>6.00</b>   | <b>702.00</b> | <b>3.00</b>   | <b>61.00</b>  | <b>1010.00</b> |                |
| Exp.         | 0.75           | 177.16        | 4.49          | 524.76        | 2.24          | 45.60         |                | <b>p-value</b> |
| Exp.         | 0.25           | 59.84         | 1.51          | 177.24        | 0.76          | 15.40         |                | 0.220444513*   |
| X^2          | 0.09           | 0.03          | 1.38          | 0.02          | 0.26          | 0.00          | <b>1.77</b>    |                |
| X^2          | 0.25           | 0.08          | 4.08          | 0.06          | 0.76          | 0.01          | <b>5.23</b>    |                |
| CTG          | 462.00         | 1.00          | 34.00         | 2.00          | 0.00          | 0.00          | 499.00         |                |
| CTG          | 144.00         | 0.00          | 2.00          | 0.00          | 0.00          | 0.00          | 146.00         |                |
| <b>TOTAL</b> | <b>606.00</b>  | <b>1.00</b>   | <b>36.00</b>  | <b>2.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>645.00</b>  |                |
| Exp.         | 468.83         | 0.77          | 27.85         | 1.55          | 0.25          | 0.25          |                | <b>p-value</b> |
| Exp.         | 137.17         | 0.25          | 8.15          | 0.45          | 0.25          | 0.25          |                | 0.214183285*   |
| X^2          | 0.10           | 0.07          | 1.36          | 0.13          | 0.25          | 0.25          | <b>2.16</b>    |                |
| X^2          | 0.34           | 0.25          | 4.64          | 0.45          | 0.25          | 0.25          | <b>6.18</b>    |                |
| CAG          | 0.00           | 0.00          | 0.00          | 176.00        | 50.00         | 327.00        | 553.00         |                |
| CAG          | 0.00           | 0.00          | 1.00          | 60.00         | 27.00         | 113.00        | 201.00         |                |
| <b>TOTAL</b> | <b>0.00</b>    | <b>0.00</b>   | <b>1.00</b>   | <b>236.00</b> | <b>77.00</b>  | <b>440.00</b> | <b>754.00</b>  |                |
| Exp.         | 0.25           | 0.25          | 0.73          | 173.09        | 56.47         | 322.71        |                | <b>p-value</b> |
| Exp.         | 0.25           | 0.25          | 0.27          | 62.91         | 20.53         | 117.29        |                | 0.312788018*   |
| X^2          | 0.25           | 0.25          | 0.73          | 0.05          | 0.74          | 0.06          | <b>2.08</b>    |                |
| X^2          | 0.25           | 0.25          | 2.02          | 0.13          | 2.04          | 0.16          | <b>4.85</b>    |                |
| CGA          | 0.00           | 0.00          | 0.00          | 0.00          | 0.00          | 0.00          | 0.00           |                |
| CGA          | 0.00           | 0.00          | 0.00          | 0.00          | 0.00          | 0.00          | 0.00           |                |
| <b>TOTAL</b> | <b>0.00</b>    | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>    |                |
| Exp.         | N/A            | N/A           | N/A           | N/A           | N/A           | N/A           |                | <b>p-value</b> |
| Exp.         | N/A            | N/A           | N/A           | N/A           | N/A           | N/A           |                | N/A            |
| X^2          | N/A            | N/A           | N/A           | N/A           | N/A           | N/A           | <b>N/A</b>     |                |
| X^2          | N/A            | N/A           | N/A           | N/A           | N/A           | N/A           | <b>N/A</b>     |                |
| CGG          | 1.00           | 0.00          | 110.00        | 1.00          | 1.00          | 0.00          | 113.00         |                |
| CGG          | 1.00           | 0.00          | 38.00         | 0.00          | 2.00          | 0.00          | 41.00          |                |
| <b>TOTAL</b> | <b>2.00</b>    | <b>0.00</b>   | <b>148.00</b> | <b>1.00</b>   | <b>3.00</b>   | <b>0.00</b>   | <b>154.00</b>  |                |
| Exp.         | 1.47           | 0.25          | 108.60        | 0.73          | 2.20          | 0.25          |                | <b>p-value</b> |
| Exp.         | 0.53           | 0.25          | 39.40         | 0.27          | 0.80          | 0.25          |                | 0.630549383*   |
| X^2          | 0.15           | 0.25          | 0.02          | 0.10          | 0.66          | 0.25          | <b>1.42</b>    |                |
| X^2          | 0.41           | 0.25          | 0.05          | 0.27          | 1.81          | 0.25          | <b>3.03</b>    |                |
| CGC          | 0.00           | 1.00          | 1.00          | 96.00         | 0.00          | 16.00         | 114.00         |                |
| CGC          | 10.00          | 1.00          | 1.00          | 42.00         | 0.00          | 1.00          | 55.00          |                |
| <b>TOTAL</b> | <b>10.00</b>   | <b>2.00</b>   | <b>2.00</b>   | <b>138.00</b> | <b>0.00</b>   | <b>17.00</b>  | <b>169.00</b>  |                |
| Exp.         | 6.75           | 1.35          | 1.35          | 93.09         | 0.25          | 11.47         |                | <b>p-value</b> |
| Exp.         | 3.25           | 0.65          | 0.65          | 44.91         | 0.25          | 5.53          |                | 5.53538E-05*   |
| X^2          | 6.75           | 0.09          | 0.09          | 0.09          | 0.25          | 1.79          | <b>9.06</b>    |                |
| X^2          | 13.98          | 0.19          | 0.19          | 0.19          | 0.25          | 3.71          | <b>18.51</b>   |                |
| CGT          | 86.00          | 15.00         | 93.00         | 89.00         | 5.00          | 17.00         | 305.00         |                |
| CGT          | 47.00          | 5.00          | 5.00          | 30.00         | 0.00          | 2.00          | 89.00          |                |
| <b>TOTAL</b> | <b>133.00</b>  | <b>20.00</b>  | <b>98.00</b>  | <b>119.00</b> | <b>5.00</b>   | <b>19.00</b>  | <b>394.00</b>  |                |
| Exp.         | 102.96         | 15.48         | 75.86         | 92.12         | 3.87          | 14.71         |                | <b>p-value</b> |
| Exp.         | 30.04          | 4.52          | 22.14         | 26.88         | 1.13          | 4.29          |                | 3.63629E-06*   |
| X^2          | 2.79           | 0.02          | 3.87          | 0.11          | 0.33          | 0.36          | <b>7.47</b>    |                |
| X^2          | 9.57           | 0.05          | 13.27         | 0.36          | 1.13          | 1.22          | <b>25.60</b>   |                |
| TAA          | 6.00           | 82.00         | 6.00          | 376.00        | 16.00         | 26.00         | 512.00         |                |
| TAA          | 1.00           | 12.00         | 0.00          | 102.00        | 3.00          | 36.00         | 154.00         |                |
| <b>TOTAL</b> | <b>7.00</b>    | <b>94.00</b>  | <b>6.00</b>   | <b>478.00</b> | <b>19.00</b>  | <b>62.00</b>  | <b>666.00</b>  |                |
| Exp.         | 5.38           | 72.26         | 4.61          | 367.47        | 14.61         | 47.66         |                | <b>p-value</b> |
| Exp.         | 1.62           | 21.74         | 1.39          | 110.53        | 4.39          | 14.34         |                | 5.93505E-10*   |
| X^2          | 0.07           | 1.31          | 0.42          | 0.20          | 0.13          | 9.85          | <b>11.98</b>   |                |
| X^2          | 0.24           | 4.36          | 1.39          | 0.66          | 0.44          | 32.74         | <b>39.82</b>   |                |
| TAC          | 878.00         | 152.00        | 408.00        | 88.00         | 26.00         | 71.00         | 1623.00        |                |
| TAC          | 335.00         | 59.00         | 130.00        | 37.00         | 9.00          | 8.00          | 578.00         |                |
| <b>TOTAL</b> | <b>1213.00</b> | <b>211.00</b> | <b>538.00</b> | <b>125.00</b> | <b>35.00</b>  | <b>79.00</b>  | <b>2201.00</b> |                |
| Exp.         | 894.46         | 155.59        | 396.72        | 92.17         | 25.81         | 58.25         |                | <b>p-value</b> |
| Exp.         | 318.54         | 55.41         | 141.28        | 32.83         | 9.19          | 20.75         |                | 0.015386859    |

|              |               |               |               |               |               |               |                |                |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|----------------|
| X^2          | 0.30          | 0.08          | 0.32          | 0.19          | 0.00          | 2.79          | <b>3.69</b>    |                |
| X^2          | 0.85          | 0.23          | 0.90          | 0.53          | 0.00          | 7.83          | <b>10.35</b>   |                |
| TCA          | 9.00          | 1.00          | 210.00        | 83.00         | 608.00        | 44.00         | 955.00         |                |
| TCA          | 1.00          | 3.00          | 19.00         | 26.00         | 320.00        | 7.00          | 376.00         |                |
| <b>TOTAL</b> | <b>10.00</b>  | <b>4.00</b>   | <b>229.00</b> | <b>109.00</b> | <b>928.00</b> | <b>51.00</b>  | <b>1331.00</b> |                |
| Exp.         | 7.18          | 2.87          | 164.31        | 78.21         | 665.85        | 36.59         |                | <b>p-value</b> |
| Exp.         | 2.82          | 1.13          | 64.69         | 30.79         | 262.15        | 14.41         |                | 8.99632E-15    |
| X^2          | 0.46          | 1.22          | 12.71         | 0.29          | 5.03          | 1.50          | <b>21.21</b>   |                |
| X^2          | 1.18          | 3.09          | 32.27         | 0.75          | 12.76         | 3.81          | <b>53.86</b>   |                |
| TCC          | 28.00         | 84.00         | 175.00        | 29.00         | 259.00        | 16.00         | 591.00         |                |
| TCC          | 6.00          | 29.00         | 73.00         | 16.00         | 73.00         | 5.00          | 202.00         |                |
| <b>TOTAL</b> | <b>34.00</b>  | <b>113.00</b> | <b>248.00</b> | <b>45.00</b>  | <b>332.00</b> | <b>21.00</b>  | <b>793.00</b>  |                |
| Exp.         | 25.34         | 84.22         | 184.83        | 33.54         | 247.43        | 15.65         |                | <b>p-value</b> |
| Exp.         | 8.66          | 28.78         | 63.17         | 11.46         | 84.57         | 5.35          |                | 0.172686462    |
| X^2          | 0.28          | 0.00          | 0.52          | 0.61          | 0.54          | 0.01          | <b>1.97</b>    |                |
| X^2          | 0.82          | 0.00          | 1.53          | 1.80          | 1.58          | 0.02          | <b>5.75</b>    |                |
| TCT          | 5.00          | 0.00          | 287.00        | 6.00          | 163.00        | 1.00          | 462.00         |                |
| TCT          | 1.00          | 0.00          | 68.00         | 0.00          | 36.00         | 0.00          | 105.00         |                |
| <b>TOTAL</b> | <b>6.00</b>   | <b>0.00</b>   | <b>355.00</b> | <b>6.00</b>   | <b>199.00</b> | <b>1.00</b>   | <b>567.00</b>  |                |
| Exp.         | 4.89          | 0.25          | 289.26        | 4.89          | 162.15        | 0.81          |                | <b>p-value</b> |
| Exp.         | 1.11          | 0.25          | 65.74         | 1.11          | 36.85         | 0.25          |                | 0.908547618*   |
| X^2          | 0.00          | 0.25          | 0.02          | 0.25          | 0.00          | 0.04          | <b>0.57</b>    |                |
| X^2          | 0.01          | 0.25          | 0.08          | 1.11          | 0.02          | 0.25          | <b>1.72</b>    |                |
| TCG          | 0.00          | 0.00          | 3.00          | 1.00          | 91.00         | 96.00         | 191.00         |                |
| TCG          | 1.00          | 0.00          | 1.00          | 0.00          | 67.00         | 6.00          | 75.00          |                |
| <b>TOTAL</b> | <b>1.00</b>   | <b>0.00</b>   | <b>4.00</b>   | <b>1.00</b>   | <b>158.00</b> | <b>102.00</b> | <b>266.00</b>  |                |
| Exp.         | 0.72          | 0.25          | 2.87          | 0.72          | 113.45        | 73.24         |                | <b>p-value</b> |
| Exp.         | 0.28          | 0.25          | 1.13          | 0.28          | 44.55         | 28.76         |                | 2.5424E-08*    |
| X^2          | 0.72          | 0.25          | 0.01          | 0.11          | 4.44          | 7.07          | <b>12.60</b>   |                |
| X^2          | 1.83          | 0.25          | 0.01          | 0.28          | 11.31         | 18.01         | <b>31.70</b>   |                |
| TAT          | 63.00         | 89.00         | 201.00        | 69.00         | 198.00        | 11.00         | 631.00         |                |
| TAT          | 36.00         | 35.00         | 110.00        | 29.00         | 17.00         | 1.00          | 228.00         |                |
| <b>TOTAL</b> | <b>99.00</b>  | <b>124.00</b> | <b>311.00</b> | <b>98.00</b>  | <b>215.00</b> | <b>12.00</b>  | <b>859.00</b>  |                |
| Exp.         | 72.72         | 91.09         | 228.45        | 71.99         | 157.93        | 8.81          |                | <b>p-value</b> |
| Exp.         | 26.28         | 32.91         | 82.55         | 26.01         | 57.07         | 3.19          |                | 2.71471E-11    |
| X^2          | 1.30          | 0.05          | 3.30          | 0.12          | 10.16         | 0.54          | <b>15.48</b>   |                |
| X^2          | 3.60          | 0.13          | 9.13          | 0.34          | 28.13         | 1.50          | <b>42.83</b>   |                |
| TTA          | 153.00        | 41.00         | 221.00        | 45.00         | 41.00         | 91.00         | 592.00         |                |
| TTA          | 21.00         | 9.00          | 108.00        | 5.00          | 22.00         | 26.00         | 191.00         |                |
| <b>TOTAL</b> | <b>174.00</b> | <b>50.00</b>  | <b>329.00</b> | <b>50.00</b>  | <b>63.00</b>  | <b>117.00</b> | <b>783.00</b>  |                |
| Exp.         | 131.56        | 37.80         | 248.75        | 37.80         | 47.63         | 88.46         |                | <b>p-value</b> |
| Exp.         | 42.44         | 12.20         | 80.25         | 12.20         | 15.37         | 28.54         |                | 4.08815E-07    |
| X^2          | 3.50          | 0.27          | 3.09          | 1.37          | 0.92          | 0.07          | <b>9.23</b>    |                |
| X^2          | 10.83         | 0.84          | 9.59          | 4.25          | 2.86          | 0.23          | <b>28.60</b>   |                |
| TTT          | 111.00        | 184.00        | 5.00          | 139.00        | 1.00          | 24.00         | 464.00         |                |
| TTT          | 20.00         | 11.00         | 1.00          | 24.00         | 7.00          | 76.00         | 139.00         |                |
| <b>TOTAL</b> | <b>131.00</b> | <b>195.00</b> | <b>6.00</b>   | <b>163.00</b> | <b>8.00</b>   | <b>100.00</b> | <b>603.00</b>  |                |
| Exp.         | 100.80        | 150.05        | 4.62          | 125.43        | 6.16          | 76.95         |                | <b>p-value</b> |
| Exp.         | 30.20         | 44.95         | 1.38          | 37.57         | 1.84          | 23.05         |                | 8.64189E-46*   |
| X^2          | 1.03          | 7.68          | 0.03          | 1.47          | 4.32          | 36.43         | <b>50.97</b>   |                |
| X^2          | 3.44          | 25.64         | 0.11          | 4.90          | 14.42         | 121.62        | <b>170.13</b>  |                |
| TTC          | 3.00          | 78.00         | 0.00          | 198.00        | 2.00          | 185.00        | 466.00         |                |
| TTC          | 1.00          | 24.00         | 0.00          | 67.00         | 0.00          | 56.00         | 148.00         |                |
| <b>TOTAL</b> | <b>4.00</b>   | <b>102.00</b> | <b>0.00</b>   | <b>265.00</b> | <b>2.00</b>   | <b>241.00</b> | <b>614.00</b>  |                |
| Exp.         | 3.04          | 77.41         | 0.25          | 201.12        | 1.52          | 182.91        |                | <b>p-value</b> |
| Exp.         | 0.96          | 24.59         | 0.25          | 63.88         | 0.48          | 58.09         |                | 0.966048267*   |
| X^2          | 0.00          | 0.00          | 0.25          | 0.05          | 0.15          | 0.02          | <b>0.48</b>    |                |
| X^2          | 0.00          | 0.01          | 0.25          | 0.15          | 0.48          | 0.08          | <b>0.98</b>    |                |
| TTG          | 239.00        | 1.00          | 233.00        | 0.00          | 0.00          | 0.00          | 473.00         |                |
| TTG          | 67.00         | 0.00          | 114.00        | 0.00          | 0.00          | 0.00          | 181.00         |                |
| <b>TOTAL</b> | <b>306.00</b> | <b>1.00</b>   | <b>347.00</b> | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>654.00</b>  |                |
| Exp.         | 221.31        | 0.72          | 250.96        | 0.25          | 0.25          | 0.25          |                | <b>p-value</b> |
| Exp.         | 84.69         | 0.28          | 96.04         | 0.25          | 0.25          | 0.25          |                | 0.071437393*   |
| X^2          | 1.41          | 0.11          | 1.29          | 0.25          | 0.25          | 0.25          | <b>3.56</b>    |                |
| X^2          | 3.69          | 0.28          | 3.36          | 0.25          | 0.25          | 0.25          | <b>8.08</b>    |                |
| TAG          | 0.00          | 0.00          | 0.00          | 251.00        | 19.00         | 187.00        | 457.00         |                |
| TAG          | 0.00          | 0.00          | 0.00          | 92.00         | 14.00         | 58.00         | 164.00         |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|              |        |        |       |        |        |        |              |                |
|--------------|--------|--------|-------|--------|--------|--------|--------------|----------------|
| <b>TOTAL</b> | 0.00   | 0.00   | 0.00  | 343.00 | 33.00  | 245.00 | 621.00       |                |
| Exp.         | 0.25   | 0.25   | 0.25  | 252.42 | 24.29  | 180.30 |              | <b>p-value</b> |
| Exp.         | 0.25   | 0.25   | 0.25  | 90.58  | 8.71   | 64.70  |              | 0.377102578*   |
| X^2          | 0.25   | 0.25   | 0.25  | 0.01   | 1.15   | 0.25   | <b>2.16</b>  |                |
| X^2          | 0.25   | 0.25   | 0.25  | 0.02   | 3.20   | 0.69   | <b>4.67</b>  |                |
| <b>TGA</b>   | 65.00  | 157.00 | 3.00  | 212.00 | 0.00   | 1.00   | 438.00       |                |
| TGA          | 28.00  | 74.00  | 0.00  | 67.00  | 0.00   | 0.00   | 169.00       |                |
| <b>TOTAL</b> | 93.00  | 231.00 | 3.00  | 279.00 | 0.00   | 1.00   | 607.00       |                |
| Exp.         | 67.11  | 166.69 | 2.16  | 201.32 | 0.25   | 0.72   |              | <b>p-value</b> |
| Exp.         | 25.89  | 64.31  | 0.84  | 77.68  | 0.25   | 0.28   |              | 0.322424272*   |
| X^2          | 0.07   | 0.56   | 0.32  | 0.57   | 0.25   | 0.11   | <b>1.88</b>  |                |
| X^2          | 0.17   | 1.46   | 0.84  | 1.47   | 0.25   | 0.28   | <b>4.46</b>  |                |
| <b>TGG</b>   | 0.00   | 223.00 | 0.00  | 37.00  | 0.00   | 0.00   | 260.00       |                |
| TGG          | 0.00   | 59.00  | 0.00  | 29.00  | 0.00   | 0.00   | 88.00        |                |
| <b>TOTAL</b> | 0.00   | 282.00 | 0.00  | 66.00  | 0.00   | 0.00   | 348.00       |                |
| Exp.         | 0.25   | 210.69 | 0.25  | 49.31  | 0.25   | 0.25   |              | <b>p-value</b> |
| Exp.         | 0.25   | 71.31  | 0.25  | 16.69  | 0.25   | 0.25   |              | 0.010371514*   |
| X^2          | 0.25   | 0.72   | 0.25  | 3.07   | 0.25   | 0.25   | <b>4.79</b>  |                |
| X^2          | 0.25   | 2.13   | 0.25  | 9.08   | 0.25   | 0.25   | <b>12.21</b> |                |
| <b>TGC</b>   | 7.00   | 99.00  | 0.00  | 237.00 | 0.00   | 0.00   | 343.00       |                |
| TGC          | 4.00   | 34.00  | 0.00  | 85.00  | 0.00   | 0.00   | 123.00       |                |
| <b>TOTAL</b> | 11.00  | 133.00 | 0.00  | 322.00 | 0.00   | 0.00   | 466.00       |                |
| Exp.         | 8.10   | 97.89  | 0.25  | 237.01 | 0.25   | 0.25   |              | <b>p-value</b> |
| Exp.         | 2.90   | 35.11  | 0.25  | 84.99  | 0.25   | 0.25   |              | 0.987543905*   |
| X^2          | 0.15   | 0.01   | 0.25  | 0.00   | 0.25   | 0.25   | <b>0.91</b>  |                |
| X^2          | 0.41   | 0.03   | 0.25  | 0.00   | 0.25   | 0.25   | <b>1.20</b>  |                |
| <b>TGT</b>   | 0.00   | 17.00  | 0.00  | 387.00 | 2.00   | 1.00   | 407.00       |                |
| TGT          | 0.00   | 1.00   | 0.00  | 135.00 | 0.00   | 1.00   | 137.00       |                |
| <b>TOTAL</b> | 0.00   | 18.00  | 0.00  | 522.00 | 2.00   | 2.00   | 544.00       |                |
| Exp.         | 0.25   | 13.47  | 0.25  | 390.54 | 1.50   | 1.50   |              | <b>p-value</b> |
| Exp.         | 0.25   | 4.53   | 0.25  | 131.46 | 0.50   | 0.50   |              | 0.399630721*   |
| X^2          | 0.25   | 0.93   | 0.25  | 0.03   | 0.17   | 0.16   | <b>1.79</b>  |                |
| X^2          | 0.25   | 2.75   | 0.25  | 0.10   | 0.50   | 0.49   | <b>4.34</b>  |                |
| <b>GAA</b>   | 90.00  | 5.00   | 3.00  | 4.00   | 49.00  | 58.00  | 209.00       |                |
| GAA          | 35.00  | 0.00   | 0.00  | 0.00   | 29.00  | 10.00  | 74.00        |                |
| <b>TOTAL</b> | 125.00 | 5.00   | 3.00  | 4.00   | 78.00  | 68.00  | 283.00       |                |
| Exp.         | 92.31  | 3.69   | 2.22  | 2.95   | 57.60  | 50.22  |              | <b>p-value</b> |
| Exp.         | 32.69  | 1.31   | 0.78  | 1.05   | 20.40  | 17.78  |              | 0.015633318*   |
| X^2          | 0.06   | 0.46   | 0.28  | 0.37   | 1.29   | 1.21   | <b>3.66</b>  |                |
| X^2          | 0.16   | 1.31   | 0.78  | 1.05   | 3.63   | 3.40   | <b>10.34</b> |                |
| <b>GAC</b>   | 0.00   | 1.00   | 2.00  | 58.00  | 142.00 | 116.00 | 319.00       |                |
| GAC          | 0.00   | 1.00   | 0.00  | 26.00  | 57.00  | 41.00  | 125.00       |                |
| <b>TOTAL</b> | 0.00   | 2.00   | 2.00  | 84.00  | 199.00 | 157.00 | 444.00       |                |
| Exp.         | 0.25   | 1.44   | 1.44  | 60.35  | 142.98 | 112.80 |              | <b>p-value</b> |
| Exp.         | 0.25   | 0.56   | 0.56  | 23.65  | 56.02  | 44.20  |              | 0.859124216*   |
| X^2          | 0.25   | 0.13   | 0.22  | 0.09   | 0.01   | 0.09   | <b>0.79</b>  |                |
| X^2          | 0.25   | 0.34   | 0.56  | 0.23   | 0.02   | 0.23   | <b>1.63</b>  |                |
| <b>GCA</b>   | 142.00 | 108.00 | 37.00 | 94.00  | 90.00  | 104.00 | 575.00       |                |
| GCA          | 33.00  | 37.00  | 13.00 | 36.00  | 33.00  | 36.00  | 188.00       |                |
| <b>TOTAL</b> | 175.00 | 145.00 | 50.00 | 130.00 | 123.00 | 140.00 | 763.00       |                |
| Exp.         | 131.88 | 109.27 | 37.68 | 97.97  | 92.69  | 105.50 |              | <b>p-value</b> |
| Exp.         | 43.12  | 35.73  | 12.32 | 32.03  | 30.31  | 34.50  |              | 0.504545459    |
| X^2          | 0.78   | 0.01   | 0.01  | 0.16   | 0.08   | 0.02   | <b>1.06</b>  |                |
| X^2          | 2.37   | 0.05   | 0.04  | 0.49   | 0.24   | 0.07   | <b>3.25</b>  |                |
| <b>GCC</b>   | 209.00 | 66.00  | 31.00 | 4.00   | 115.00 | 1.00   | 426.00       |                |
| GCC          | 97.00  | 16.00  | 3.00  | 0.00   | 41.00  | 2.00   | 159.00       |                |
| <b>TOTAL</b> | 306.00 | 82.00  | 34.00 | 4.00   | 156.00 | 3.00   | 585.00       |                |
| Exp.         | 222.83 | 59.71  | 24.76 | 2.91   | 113.60 | 2.18   |              | <b>p-value</b> |
| Exp.         | 83.17  | 22.29  | 9.24  | 1.09   | 42.40  | 0.82   |              | 0.009146875*   |
| X^2          | 0.86   | 0.66   | 1.57  | 0.41   | 0.02   | 0.64   | <b>4.16</b>  |                |
| X^2          | 2.30   | 1.77   | 4.21  | 1.09   | 0.05   | 1.72   | <b>11.14</b> |                |
| <b>GCT</b>   | 0.00   | 0.00   | 32.00 | 7.00   | 78.00  | 100.00 | 217.00       |                |
| GCT          | 0.00   | 0.00   | 1.00  | 4.00   | 40.00  | 34.00  | 79.00        |                |
| <b>TOTAL</b> | 0.00   | 0.00   | 33.00 | 11.00  | 118.00 | 134.00 | 296.00       |                |
| Exp.         | 0.25   | 0.25   | 24.19 | 8.06   | 86.51  | 98.24  |              | <b>p-value</b> |
| Exp.         | 0.25   | 0.25   | 8.81  | 2.94   | 31.49  | 35.76  |              | 0.021404569*   |
| X^2          | 0.25   | 0.25   | 2.52  | 0.14   | 0.84   | 0.03   | <b>4.03</b>  |                |
| X^2          | 0.25   | 0.25   | 6.92  | 0.39   | 2.30   | 0.09   | <b>10.19</b> |                |

|              |               |               |              |               |               |               |                |             |                |
|--------------|---------------|---------------|--------------|---------------|---------------|---------------|----------------|-------------|----------------|
| GCG          | 0.00          | 0.00          | 0.00         | 0.00          | 0.00          | 0.00          | 0.00           | 0.00        |                |
| GCG          | 0.00          | 0.00          | 0.00         | 0.00          | 0.00          | 0.00          | 0.00           | 0.00        |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>  | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>    | <b>0.00</b> |                |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | N/A            |             | <b>p-value</b> |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | N/A            |             | N/A            |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | N/A            | <b>N/A</b>  |                |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | N/A            | <b>N/A</b>  |                |
| GAT          | 0.00          | 145.00        | 0.00         | 147.00        | 0.00          | 17.00         | 309.00         |             |                |
| GAT          | 0.00          | 63.00         | 0.00         | 54.00         | 0.00          | 6.00          | 123.00         |             |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>208.00</b> | <b>0.00</b>  | <b>201.00</b> | <b>0.00</b>   | <b>23.00</b>  | <b>432.00</b>  |             |                |
| Exp.         | 0.25          | 148.78        | 0.25         | 143.77        | 0.25          | 16.45         |                |             | <b>p-value</b> |
| Exp.         | 0.25          | 59.22         | 0.25         | 57.23         | 0.25          | 6.55          |                |             | 0.985301073*   |
| X^2          | 0.25          | 0.10          | 0.25         | 0.07          | 0.25          | 0.02          | <b>0.94</b>    |             |                |
| X^2          | 0.25          | 0.24          | 0.25         | 0.18          | 0.25          | 0.05          | <b>1.22</b>    |             |                |
| GTA          | 329.00        | 173.00        | 49.00        | 176.00        | 266.00        | 229.00        | 1222.00        |             |                |
| GTA          | 124.00        | 43.00         | 24.00        | 61.00         | 98.00         | 92.00         | 442.00         |             |                |
| <b>TOTAL</b> | <b>453.00</b> | <b>216.00</b> | <b>73.00</b> | <b>237.00</b> | <b>364.00</b> | <b>321.00</b> | <b>1664.00</b> |             |                |
| Exp.         | 332.67        | 158.63        | 53.61        | 174.05        | 267.31        | 235.73        |                |             | <b>p-value</b> |
| Exp.         | 120.33        | 57.38         | 19.39        | 62.95         | 96.69         | 85.27         |                |             | 0.19388313     |
| X^2          | 0.04          | 1.30          | 0.40         | 0.02          | 0.01          | 0.19          | <b>1.96</b>    |             |                |
| X^2          | 0.11          | 3.60          | 1.10         | 0.06          | 0.02          | 0.53          | <b>5.42</b>    |             |                |
| GTT          | 0.00          | 0.00          | 0.00         | 0.00          | 90.00         | 1.00          | 91.00          |             |                |
| GTT          | 0.00          | 0.00          | 0.00         | 0.00          | 22.00         | 0.00          | 22.00          |             |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>  | <b>0.00</b>   | <b>112.00</b> | <b>1.00</b>   | <b>113.00</b>  |             |                |
| Exp.         | 0.25          | 0.25          | 0.25         | 0.25          | 90.19         | 0.81          |                |             | <b>p-value</b> |
| Exp.         | 0.25          | 0.25          | 0.25         | 0.25          | 21.81         | 0.25          |                |             | 0.999971899*   |
| X^2          | 0.25          | 0.25          | 0.25         | 0.25          | 0.00          | 0.05          | <b>1.05</b>    |             |                |
| X^2          | 0.25          | 0.25          | 0.25         | 0.25          | 0.00          | 0.25          | <b>1.25</b>    |             |                |
| GTC          | 85.00         | 73.00         | 1.00         | 12.00         | 0.00          | 37.00         | 208.00         |             |                |
| GTC          | 22.00         | 10.00         | 9.00         | 4.00          | 3.00          | 24.00         | 72.00          |             |                |
| <b>TOTAL</b> | <b>107.00</b> | <b>83.00</b>  | <b>10.00</b> | <b>16.00</b>  | <b>3.00</b>   | <b>61.00</b>  | <b>280.00</b>  |             |                |
| Exp.         | 79.49         | 61.66         | 7.43         | 11.89         | 2.23          | 45.31         |                |             | <b>p-value</b> |
| Exp.         | 27.51         | 21.34         | 2.57         | 4.11          | 0.77          | 15.69         |                |             | 9.78592E-09*   |
| X^2          | 0.38          | 2.09          | 5.56         | 0.00          | 2.23          | 1.53          | <b>11.79</b>   |             |                |
| X^2          | 1.11          | 6.03          | 16.07        | 0.00          | 6.44          | 4.41          | <b>34.05</b>   |             |                |
| GTG          | 1.00          | 0.00          | 5.00         | 1.00          | 2.00          | 3.00          | 12.00          |             |                |
| GTG          | 0.00          | 0.00          | 0.00         | 0.00          | 0.00          | 0.00          | 0.00           |             |                |
| <b>TOTAL</b> | <b>1.00</b>   | <b>0.00</b>   | <b>5.00</b>  | <b>1.00</b>   | <b>2.00</b>   | <b>3.00</b>   | <b>12.00</b>   |             |                |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           |                |             | <b>p-value</b> |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           |                |             | N/A            |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | <b>N/A</b>     |             |                |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | <b>N/A</b>     |             |                |
| GAG          | 0.00          | 0.00          | 0.00         | 5.00          | 0.00          | 0.00          | 5.00           |             |                |
| GAG          | 0.00          | 0.00          | 0.00         | 0.00          | 0.00          | 0.00          | 0.00           |             |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>  | <b>5.00</b>   | <b>0.00</b>   | <b>0.00</b>   | <b>5.00</b>    |             |                |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           |                |             | <b>p-value</b> |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           |                |             | N/A            |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | <b>N/A</b>     |             |                |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           | <b>N/A</b>     |             |                |
| GGA          | 193.00        | 4.00          | 1.00         | 0.00          | 48.00         | 13.00         | 259.00         |             |                |
| GGA          | 66.00         | 1.00          | 0.00         | 0.00          | 22.00         | 5.00          | 94.00          |             |                |
| <b>TOTAL</b> | <b>259.00</b> | <b>5.00</b>   | <b>1.00</b>  | <b>0.00</b>   | <b>70.00</b>  | <b>18.00</b>  | <b>353.00</b>  |             |                |
| Exp.         | 190.03        | 3.67          | 0.73         | 0.25          | 51.36         | 13.21         |                |             | <b>p-value</b> |
| Exp.         | 68.97         | 1.33          | 0.27         | 0.25          | 18.64         | 4.79          |                |             | 0.914555051*   |
| X^2          | 0.05          | 0.03          | 0.10         | 0.25          | 0.22          | 0.00          | <b>0.65</b>    |             |                |
| X^2          | 0.13          | 0.08          | 0.27         | 0.25          | 0.61          | 0.01          | <b>1.34</b>    |             |                |
| GGG          | 14.00         | 29.00         | 0.00         | 1.00          | 0.00          | 197.00        | 241.00         |             |                |
| GGG          | 1.00          | 30.00         | 0.00         | 0.00          | 0.00          | 62.00         | 93.00          |             |                |
| <b>TOTAL</b> | <b>15.00</b>  | <b>59.00</b>  | <b>0.00</b>  | <b>1.00</b>   | <b>0.00</b>   | <b>259.00</b> | <b>334.00</b>  |             |                |
| Exp.         | 10.82         | 42.57         | 0.25         | 0.72          | 0.25          | 186.88        |                |             | <b>p-value</b> |
| Exp.         | 4.18          | 16.43         | 0.25         | 0.28          | 0.25          | 72.12         |                |             | 0.000729675*   |
| X^2          | 0.93          | 4.33          | 0.25         | 0.11          | 0.25          | 0.55          | <b>6.41</b>    |             |                |
| X^2          | 2.42          | 11.21         | 0.25         | 0.28          | 0.25          | 1.42          | <b>15.83</b>   |             |                |
| GGC          | 0.00          | 0.00          | 0.00         | 0.00          | 2.00          | 0.00          | 2.00           |             |                |
| GGC          | 0.00          | 0.00          | 0.00         | 0.00          | 0.00          | 0.00          | 0.00           |             |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>0.00</b>   | <b>0.00</b>  | <b>0.00</b>   | <b>2.00</b>   | <b>0.00</b>   | <b>2.00</b>    |             |                |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A           |                |             | <b>p-value</b> |

Characterizing patterns in DNA sequence trace data through informatics tools

|                |             |             |              |               |               |              |               |                |
|----------------|-------------|-------------|--------------|---------------|---------------|--------------|---------------|----------------|
| Exp.           | N/A         | N/A         | N/A          | N/A           | N/A           | N/A          |               | N/A            |
| X <sup>2</sup> | N/A         | N/A         | N/A          | N/A           | N/A           | N/A          | N/A           |                |
| X <sup>2</sup> | N/A         | N/A         | N/A          | N/A           | N/A           | N/A          | N/A           |                |
| GGT            | 1.00        | 1.00        | 15.00        | 111.00        | 91.00         | 40.00        | 259.00        |                |
| GGT            | 0.00        | 0.00        | 2.00         | 41.00         | 21.00         | 39.00        | 103.00        |                |
| <b>TOTAL</b>   | <b>1.00</b> | <b>1.00</b> | <b>17.00</b> | <b>152.00</b> | <b>112.00</b> | <b>79.00</b> | <b>362.00</b> |                |
| Exp.           | 0.72        | 0.72        | 12.16        | 108.75        | 80.13         | 56.52        |               | <b>p-value</b> |
| Exp.           | 0.28        | 0.28        | 4.84         | 43.25         | 31.87         | 22.48        |               | 0.000114644*   |
| X <sup>2</sup> | 0.11        | 0.11        | 0.66         | 0.05          | 1.47          | 4.83         | <b>7.24</b>   |                |
| X <sup>2</sup> | 0.28        | 0.28        | 1.66         | 0.12          | 3.71          | 12.14        | <b>18.20</b>  |                |

\*Kolmogorov-Smirnov test performed and generally consistent with the Chi Square test.

**Table B.** The p-values for the comparison between Laboratory A and Laboratory B for each frame and each of the primers are provided. Any p-value < 7.8125E-4 (0.05/64) is considered statistically significant with an overall  $\alpha = 0.05$  for that frame given the null hypothesis that the distribution of patterns is the same. All p-values exceeding 7.8125E-4 are black reversed; the null hypothesis is not rejected. All N/A correspond to no observations for that frame.

| Frame | p-value      |              |              |              |
|-------|--------------|--------------|--------------|--------------|
|       | A1           | B1           | C1           | D1           |
| AAA   | 0.805545483  | 0.021028387* | 0.15527684   | 0.041154751  |
| AAAC  | 0.001683344  | 0.853187729* | 7.8697E-08   | 0.173902987* |
| ACA   | 0.000246578  | 0.190484114* | 0.37142707   | 0.211463948* |
| ACC   | 3.28633E-05  | 0.056276928* | 5.38378E-11  | 0.787496426* |
| ACT   | 0.000552211  | 0.048671752* | 0.349082834  | 0.915116185* |
| ACG   | 0.385636227* | 0.914056021* | 0.667375479* | 0.104641466* |
| AAT   | 0.001045754  | 0.001146891  | 0.707583754  | 0.193459763  |
| ATA   | 8.30891E-06  | 0.072346356  | 0.000309195* | 1.18322E-07  |
| ATT   | 1.97104E-48  | 3.44426E-13  | 3.5702E-22   | 2.28456E-06  |
| ATC   | 0.000665719  | 0.014231848* | 3.38954E-09* | 0.000634051* |
| ATG   | 0.624053924* | 0.00031568   | 0.798578346* | 1.46476E-06  |
| AAG   | 9.05752E-07  | 0.046995814* | 1.44323E-05* | 0.56795021*  |
| AGA   | 0.994592443* | 0.181656286* | 0.029574993* | 0.774247395* |
| AGG   | 0.311414387* | 3.13192E-05* | 0.897599027* | 0.918656398* |
| AGC   | 0.057880438  | 0.799939444* | 0.178391345* | 0.634620207* |
| AGT   | 1.05449E-12* | 0.007445644* | 6.34677E-14* | 0.800142041* |
| CAA   | 0.582330595  | 0.008446692* | 0.813047819* | 0.988619265* |
| CAC   | 0.387273169  | N/A          | 0.110483163  | 0.424859978* |
| CCA   | 0.003573017  | 3.82091E-05  | 8.09899E-12  | 0.921384507* |
| CCC   | 0.002298114  | 0.287932574* | 0.342186048* | 0.986800968* |
| CCT   | 0.008383757  | 2.43351E-08* | 5.5433E-05   | 0.021033552* |
| CCG   | 0.040836662  | 0.99929933*  | 0.61175165   | 0.085796601* |
| CAT   | 1.25314E-07  | 0.003129742* | 1.25507E-08  | 0.983603337* |
| CTA   | 0.003385636* | 5.24768E-06  | 1.57037E-06  | 0.203329534* |
| CTT   | 0.008916897* | 0.001110402* | 1.37616E-10  | 5.0813E-06*  |
| CTC   | 0.220444513* | 0.162783656* | 0.163109874  | 0.063484446* |
| CTG   | 0.214183285* | 0.053631141* | 0.626347062* | 3.58232E-06* |
| CAG   | 0.312788018* | 0.781062*    | 0.79481545   | 0.254238506* |
| CGA   | N/A          | 0.142909021* | 0.039815033* | 0.817066519* |
| CGG   | 0.630549383* | 0.00139454*  | 3.89586E-15* | 0.850444487* |
| CGC   | 5.53538E-05* | N/A          | 0.457782718  | 0.857987337* |
| CGT   | 3.63629E-06* | 0.749491785* | 0.000302216* | 0.000155967* |
| TAA   | 5.93505E-10* | 0.297752907  | 0.003310072  | 0.297105881  |
| TAC   | 0.015386859  | 1.5669E-05   | 0.001140743  | 0.966998964* |
| TCA   | 8.99632E-15  | 0.727390027* | 2.93957E-08* | 0.96389772*  |
| TCC   | 0.172686462  | 1.68345E-06* | 7.64322E-17  | 0.044870703* |
| TCT   | 0.908547618* | 0.00299369*  | 3.80678E-10  | 1.50691E-09* |
| TCG   | 2.5424E-08*  | N/A          | 0.736008048* | 0.836821379* |
| TAT   | 2.71471E-11  | 3.07764E-05  | 1.52526E-17  | 2.03963E-11  |
| TTA   | 4.08815E-07  | 0.082337003  | 0.020072433  | 0.291726754  |
| TTT   | 8.64189E-46* | 0.000881087* | 2.50269E-36  | 0.07428074*  |
| TTC   | 0.966048267* | 1.2737E-05*  | 2.12592E-27  | 0.047191801* |
| TTG   | 0.071437393* | 0.002322039  | 0.000666125* | 0.002146119* |
| TAG   | 0.377102578* | 0.108353468* | 0.000398975* | 0.999233686* |
| TGA   | 0.322424272* | 0.087144226  | 0.099409151* | 0.192119423* |
| TGG   | 0.010371514* | 2.17339E-06  | 5.88995E-05* | 0.435826546* |
| TGC   | 0.987543905* | 0.001843282  | 0.62181951*  | 0.234658592* |
| TGT   | 0.399630721* | 1.89014E-06  | 0.01555385*  | 3.82337E-06* |
| GAA   | 0.015633318* | 0.264539721* | 0.000577676* | 0.990734443* |
| GAC   | 0.859124216* | 0.036105683* | 0.855986024* | 0.99986273*  |
| GCA   | 0.504545459  | 0.236261123  | 0.38989923   | 0.242388731  |
| GCC   | 0.009146875* | N/A          | 0.128223065  | 0.35637003*  |
| GCT   | 0.021404569* | 0.011618732  | 0.407221429  | 3.46422E-10  |
| GCG   | N/A          | 0.255054645* | 0.211557491  | 0.637831307  |
| GAT   | 0.985301073* | 0.008431232  | 0.411397128* | 0.007719256* |
| GTA   | 0.19388313   | 0.009689778  | 1.17453E-28  | 0.000158686* |
| GTT   | 0.999971899* | 0.080287238  | 0.00187769*  | 1.43503E-11  |
| GTC   | 9.78592E-09* | 0.000734651* | 0.416546495* | 0.000952275* |
| GTG   | N/A          | 0.008175723  | 4.60639E-05* | 0.001591149* |
| GAG   | N/A          | 0.001022539* | 3.30067E-09* | 0.104400885* |
| GGA   | 0.914555051* | 0.003192933  | 1.44908E-05  | 0.000128548  |
| GGG   | 0.000729675* | 0.224596561* | 0.816549782* | 0.00248417*  |
| GGC   | N/A          | 0.062701476* | 0.171480285* | 2.51252E-05  |
| GGT   | 0.000114644* | 0.203993288  | 0.180395055* | 1.02674E-10  |

\*Kolmogorov-Smirnov test performed and generally consistent with the Chi Square test.

## Appendix IV — Comparison of Frames to Different Primers

The following table is an example of the processed data using the bioinformatics tools to statistically evaluate each of the 64 frames using a 2 x 6 Chi Square analysis comparing the results from data from different primers produced in the same laboratory. These results demonstrate that the pattern distributions obtained from the different primers are not equivalent. Comparisons were made between Primers A1 and C1; Primers B1 and D1; and Primers A1 and B1.

**Table A.** 2 x 6 Chi-square analysis comparing the results from Primer A1 to Primer C1.

|                | A             | B             | C             | D             | E             | F             | TOTAL         |                |
|----------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| AAA            | 71.00         | 31.00         | 76.00         | 0.00          | 58.00         | 29.00         | 265.00        |                |
| AAA            | 31.00         | 5.00          | 94.00         | 2.00          | 21.00         | 25.00         | 178.00        |                |
| <b>TOTAL</b>   | <b>102.00</b> | <b>36.00</b>  | <b>170.00</b> | <b>2.00</b>   | <b>79.00</b>  | <b>54.00</b>  | <b>443.00</b> |                |
| Exp.           | 61.02         | 21.53         | 101.69        | 1.20          | 47.26         | 32.30         |               | <b>p-value</b> |
| Exp.           | 40.98         | 14.47         | 68.31         | 0.80          | 31.74         | 21.70         |               | 1.20022E-07    |
| X <sup>2</sup> | 1.63          | 4.16          | 6.49          | 1.20          | 2.44          | 0.34          | <b>16.26</b>  |                |
| X <sup>2</sup> | 2.43          | 6.19          | 9.66          | 1.78          | 3.64          | 0.50          | <b>24.21</b>  |                |
| AAC            | 92.00         | 74.00         | 152.00        | 15.00         | 76.00         | 11.00         | 420.00        |                |
| AAC            | 116.00        | 1.00          | 26.00         | 0.00          | 87.00         | 4.00          | 234.00        |                |
| <b>TOTAL</b>   | <b>208.00</b> | <b>75.00</b>  | <b>178.00</b> | <b>15.00</b>  | <b>163.00</b> | <b>15.00</b>  | <b>654.00</b> |                |
| Exp.           | 133.58        | 48.17         | 114.31        | 9.63          | 104.68        | 9.63          |               | <b>p-value</b> |
| Exp.           | 74.42         | 26.83         | 63.69         | 5.37          | 58.32         | 5.37          |               | 1.40989E-28    |
| X <sup>2</sup> | 12.94         | 13.86         | 12.43         | 2.99          | 7.86          | 0.19          | <b>50.27</b>  |                |
| X <sup>2</sup> | 23.23         | 24.87         | 22.30         | 5.37          | 14.10         | 0.35          | <b>90.22</b>  |                |
| ACA            | 87.00         | 134.00        | 114.00        | 131.00        | 8.00          | 55.00         | 529.00        |                |
| ACA            | 31.00         | 166.00        | 1.00          | 122.00        | 32.00         | 27.00         | 379.00        |                |
| <b>TOTAL</b>   | <b>118.00</b> | <b>300.00</b> | <b>115.00</b> | <b>253.00</b> | <b>40.00</b>  | <b>82.00</b>  | <b>908.00</b> |                |
| Exp.           | 68.75         | 174.78        | 67.00         | 147.40        | 23.30         | 47.77         |               | <b>p-value</b> |
| Exp.           | 49.25         | 125.22        | 48.00         | 105.60        | 16.70         | 34.23         |               | 2.00702E-29    |
| X <sup>2</sup> | 4.85          | 9.51          | 32.97         | 1.82          | 10.05         | 1.09          | <b>60.30</b>  |                |
| X <sup>2</sup> | 6.76          | 13.28         | 46.02         | 2.55          | 14.03         | 1.53          | <b>84.17</b>  |                |
| ACC            | 126.00        | 142.00        | 153.00        | 81.00         | 128.00        | 77.00         | 707.00        |                |
| ACC            | 77.00         | 41.00         | 13.00         | 28.00         | 8.00          | 34.00         | 201.00        |                |
| <b>TOTAL</b>   | <b>203.00</b> | <b>183.00</b> | <b>166.00</b> | <b>109.00</b> | <b>136.00</b> | <b>111.00</b> | <b>908.00</b> |                |
| Exp.           | 158.06        | 142.49        | 129.25        | 84.87         | 105.89        | 86.43         |               | <b>p-value</b> |
| Exp.           | 44.94         | 40.51         | 36.75         | 24.13         | 30.11         | 24.57         |               | 7.72392E-15    |
| X <sup>2</sup> | 6.50          | 0.00          | 4.36          | 0.18          | 4.61          | 1.03          | <b>16.69</b>  |                |
| X <sup>2</sup> | 22.88         | 0.01          | 15.35         | 0.62          | 16.23         | 3.62          | <b>58.70</b>  |                |
| ACT            | 172.00        | 26.00         | 8.00          | 37.00         | 42.00         | 0.00          | 285.00        |                |
| ACT            | 13.00         | 14.00         | 18.00         | 49.00         | 0.00          | 4.00          | 98.00         |                |
| <b>TOTAL</b>   | <b>185.00</b> | <b>40.00</b>  | <b>26.00</b>  | <b>86.00</b>  | <b>42.00</b>  | <b>4.00</b>   | <b>383.00</b> |                |
| Exp.           | 137.66        | 29.77         | 19.35         | 63.99         | 31.25         | 2.98          |               | <b>p-value</b> |
| Exp.           | 47.34         | 10.23         | 6.65          | 22.01         | 10.75         | 1.02          |               | 9.3124E-27     |
| X <sup>2</sup> | 8.56          | 0.48          | 6.66          | 11.39         | 3.70          | 2.98          | <b>33.75</b>  |                |
| X <sup>2</sup> | 24.91         | 1.38          | 19.35         | 33.12         | 10.75         | 8.66          | <b>98.17</b>  |                |
| ACG            | 1.00          | 40.00         | 0.00          | 0.00          | 0.00          | 0.00          | 41.00         |                |
| ACG            | 0.00          | 1.00          | 0.00          | 60.00         | 0.00          | 27.00         | 88.00         |                |
| <b>TOTAL</b>   | <b>1.00</b>   | <b>41.00</b>  | <b>0.00</b>   | <b>60.00</b>  | <b>0.00</b>   | <b>27.00</b>  | <b>129.00</b> |                |
| Exp.           | 0.32          | 13.03         | 0.25          | 19.07         | 0.25          | 8.58          |               | <b>p-value</b> |
| Exp.           | 0.68          | 27.97         | 0.25          | 40.93         | 0.25          | 18.42         |               | 3.4923E-25*    |
| X <sup>2</sup> | 1.46          | 55.82         | 0.25          | 19.07         | 0.25          | 8.58          | <b>85.43</b>  |                |
| X <sup>2</sup> | 0.68          | 26.00         | 0.25          | 8.88          | 0.25          | 4.00          | <b>40.07</b>  |                |
| AAT            | 5.00          | 42.00         | 46.00         | 31.00         | 54.00         | 16.00         | 194.00        |                |
| AAT            | 8.00          | 40.00         | 2.00          | 32.00         | 81.00         | 140.00        | 303.00        |                |



|              |        |        |        |        |        |        |               |                |
|--------------|--------|--------|--------|--------|--------|--------|---------------|----------------|
| <b>TOTAL</b> | 13.00  | 82.00  | 48.00  | 63.00  | 135.00 | 156.00 | 497.00        |                |
| Exp.         | 5.07   | 32.01  | 18.74  | 24.59  | 52.70  | 60.89  |               | <b>p-value</b> |
| Exp.         | 7.93   | 49.99  | 29.26  | 38.41  | 82.30  | 95.11  |               | 9.02911E-26    |
| X^2          | 0.00   | 3.12   | 39.67  | 1.67   | 0.03   | 33.10  | <b>77.59</b>  |                |
| X^2          | 0.00   | 2.00   | 25.40  | 1.07   | 0.02   | 21.19  | <b>49.68</b>  |                |
| <b>ATA</b>   | 68.00  | 34.00  | 49.00  | 5.00   | 139.00 | 27.00  | 322.00        |                |
| ATA          | 70.00  | 10.00  | 71.00  | 11.00  | 34.00  | 22.00  | 218.00        |                |
| <b>TOTAL</b> | 138.00 | 44.00  | 120.00 | 16.00  | 173.00 | 49.00  | 540.00        |                |
| Exp.         | 82.29  | 26.24  | 71.56  | 9.54   | 103.16 | 29.22  |               | <b>p-value</b> |
| Exp.         | 55.71  | 17.76  | 48.44  | 6.46   | 69.84  | 19.78  |               | 6.74572E-13    |
| X^2          | 2.48   | 2.30   | 7.11   | 2.16   | 12.45  | 0.17   | <b>26.67</b>  |                |
| X^2          | 3.66   | 3.39   | 10.50  | 3.19   | 18.39  | 0.25   | <b>39.39</b>  |                |
| <b>ATT</b>   | 23.00  | 99.00  | 6.00   | 32.00  | 7.00   | 113.00 | 280.00        |                |
| ATT          | 30.00  | 10.00  | 53.00  | 4.00   | 156.00 | 96.00  | 349.00        |                |
| <b>TOTAL</b> | 53.00  | 109.00 | 59.00  | 36.00  | 163.00 | 209.00 | 629.00        |                |
| Exp.         | 23.59  | 48.52  | 26.26  | 16.03  | 72.56  | 93.04  |               | <b>p-value</b> |
| Exp.         | 29.41  | 60.48  | 32.74  | 19.97  | 90.44  | 115.96 |               | 1.99252E-55    |
| X^2          | 0.01   | 52.51  | 15.63  | 15.92  | 59.23  | 4.28   | <b>147.61</b> |                |
| X^2          | 0.01   | 42.13  | 12.54  | 12.78  | 47.52  | 3.44   | <b>118.42</b> |                |
| <b>ATC</b>   | 38.00  | 74.00  | 7.00   | 118.00 | 0.00   | 56.00  | 293.00        |                |
| ATC          | 1.00   | 6.00   | 0.00   | 81.00  | 0.00   | 62.00  | 150.00        |                |
| <b>TOTAL</b> | 39.00  | 80.00  | 7.00   | 199.00 | 0.00   | 118.00 | 443.00        |                |
| Exp.         | 25.79  | 52.91  | 4.63   | 131.62 | 0.25   | 78.05  |               | <b>p-value</b> |
| Exp.         | 13.21  | 27.09  | 2.37   | 67.38  | 0.25   | 39.95  |               | 2.65293E-13    |
| X^2          | 5.78   | 8.40   | 1.21   | 1.41   | 0.25   | 6.23   | <b>23.28</b>  |                |
| X^2          | 11.28  | 16.42  | 2.37   | 2.75   | 0.25   | 12.16  | <b>45.23</b>  |                |
| <b>ATG</b>   | 86.00  | 1.00   | 102.00 | 0.00   | 2.00   | 0.00   | 191.00        |                |
| ATG          | 79.00  | 1.00   | 35.00  | 0.00   | 0.00   | 0.00   | 115.00        |                |
| <b>TOTAL</b> | 165.00 | 2.00   | 137.00 | 0.00   | 2.00   | 0.00   | 306.00        |                |
| Exp.         | 102.99 | 1.25   | 85.51  | 0.25   | 1.25   | 0.25   |               | <b>p-value</b> |
| Exp.         | 62.01  | 0.75   | 51.49  | 0.25   | 0.75   | 0.25   |               | 0.004046262*   |
| X^2          | 2.80   | 0.05   | 3.18   | 0.25   | 0.45   | 0.25   | <b>6.98</b>   |                |
| X^2          | 4.66   | 0.08   | 5.28   | 0.25   | 0.75   | 0.25   | <b>11.27</b>  |                |
| <b>AAG</b>   | 25.00  | 63.00  | 6.00   | 75.00  | 44.00  | 2.00   | 215.00        |                |
| AAG          | 8.00   | 4.00   | 13.00  | 17.00  | 4.00   | 8.00   | 54.00         |                |
| <b>TOTAL</b> | 33.00  | 67.00  | 19.00  | 92.00  | 48.00  | 10.00  | 269.00        |                |
| Exp.         | 26.38  | 53.55  | 15.19  | 73.53  | 38.36  | 7.99   |               | <b>p-value</b> |
| Exp.         | 6.62   | 13.45  | 3.81   | 18.47  | 9.64   | 2.01   |               | 2.91673E-12    |
| X^2          | 0.07   | 1.67   | 5.56   | 0.03   | 0.83   | 4.49   | <b>12.65</b>  |                |
| X^2          | 0.29   | 6.64   | 22.12  | 0.12   | 3.30   | 17.89  | <b>50.35</b>  |                |
| <b>AGA</b>   | 0.00   | 0.00   | 61.00  | 0.00   | 0.00   | 0.00   | 61.00         |                |
| AGA          | 9.00   | 0.00   | 91.00  | 0.00   | 0.00   | 2.00   | 102.00        |                |
| <b>TOTAL</b> | 9.00   | 0.00   | 152.00 | 0.00   | 0.00   | 2.00   | 163.00        |                |
| Exp.         | 3.37   | 0.25   | 56.88  | 0.25   | 0.25   | 0.75   |               | <b>p-value</b> |
| Exp.         | 5.63   | 0.25   | 95.12  | 0.25   | 0.25   | 1.25   |               | 0.216618383*   |
| X^2          | 3.37   | 0.25   | 0.30   | 0.25   | 0.25   | 0.75   | <b>5.16</b>   |                |
| X^2          | 2.01   | 0.25   | 0.18   | 0.25   | 0.25   | 0.45   | <b>3.39</b>   |                |
| <b>AGG</b>   | 1.00   | 1.00   | 24.00  | 0.00   | 37.00  | 0.00   | 63.00         |                |
| AGG          | 9.00   | 0.00   | 49.00  | 0.00   | 0.00   | 0.00   | 58.00         |                |
| <b>TOTAL</b> | 10.00  | 1.00   | 73.00  | 0.00   | 37.00  | 0.00   | 121.00        |                |
| Exp.         | 5.21   | 0.52   | 38.01  | 0.25   | 19.26  | 0.25   |               | <b>p-value</b> |
| Exp.         | 4.79   | 0.48   | 34.99  | 0.25   | 17.74  | 0.25   |               | 3.61857E-10*   |
| X^2          | 3.40   | 0.44   | 5.16   | 0.25   | 16.33  | 0.25   | <b>25.83</b>  |                |
| X^2          | 3.69   | 0.48   | 5.61   | 0.25   | 17.74  | 0.25   | <b>28.01</b>  |                |
| <b>AGC</b>   | 47.00  | 3.00   | 91.00  | 6.00   | 71.00  | 32.00  | 250.00        |                |
| AGC          | 0.00   | 0.00   | 84.00  | 2.00   | 37.00  | 42.00  | 165.00        |                |
| <b>TOTAL</b> | 47.00  | 3.00   | 175.00 | 8.00   | 108.00 | 74.00  | 415.00        |                |
| Exp.         | 28.31  | 1.81   | 105.42 | 4.82   | 65.06  | 44.58  |               | <b>p-value</b> |
| Exp.         | 18.69  | 1.19   | 69.58  | 3.18   | 42.94  | 29.42  |               | 2.24005E-09*   |
| X^2          | 12.33  | 0.79   | 1.97   | 0.29   | 0.54   | 3.55   | <b>19.47</b>  |                |
| X^2          | 18.69  | 1.19   | 2.99   | 0.44   | 0.82   | 5.38   | <b>29.51</b>  |                |
| <b>AGT</b>   | 56.00  | 4.00   | 28.00  | 1.00   | 105.00 | 15.00  | 209.00        |                |
| AGT          | 2.00   | 10.00  | 6.00   | 12.00  | 37.00  | 1.00   | 68.00         |                |
| <b>TOTAL</b> | 58.00  | 14.00  | 34.00  | 13.00  | 142.00 | 16.00  | 277.00        |                |
| Exp.         | 43.76  | 10.56  | 25.65  | 9.81   | 107.14 | 12.07  |               | <b>p-value</b> |
| Exp.         | 14.24  | 3.44   | 8.35   | 3.19   | 34.86  | 3.93   |               | 4.9304E-13     |
| X^2          | 3.42   | 4.08   | 0.21   | 7.91   | 0.04   | 0.71   | <b>16.38</b>  |                |
| X^2          | 10.52  | 12.53  | 0.66   | 24.31  | 0.13   | 2.18   | <b>50.34</b>  |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|              |               |               |               |               |               |               |               |                |
|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|----------------|
| CAA          | 29.00         | 110.00        | 72.00         | 152.00        | 202.00        | 31.00         | 596.00        |                |
| CAA          | 0.00          | 51.00         | 0.00          | 13.00         | 68.00         | 64.00         | 196.00        |                |
| <b>TOTAL</b> | <b>29.00</b>  | <b>161.00</b> | <b>72.00</b>  | <b>165.00</b> | <b>270.00</b> | <b>95.00</b>  | <b>792.00</b> |                |
| Exp.         | 21.82         | 121.16        | 54.18         | 124.17        | 203.18        | 71.49         |               | <b>p-value</b> |
| Exp.         | 7.18          | 39.84         | 17.82         | 40.83         | 66.82         | 23.51         |               | 1.00751E-31    |
| X^2          | 2.36          | 1.03          | 5.86          | 6.24          | 0.01          | 22.93         | <b>38.43</b>  |                |
| X^2          | 7.18          | 3.12          | 17.82         | 18.97         | 0.02          | 69.73         | <b>116.84</b> |                |
| CAC          | 50.00         | 143.00        | 5.00          | 92.00         | 75.00         | 80.00         | 445.00        |                |
| CAC          | 49.00         | 45.00         | 85.00         | 0.00          | 98.00         | 24.00         | 301.00        |                |
| <b>TOTAL</b> | <b>99.00</b>  | <b>188.00</b> | <b>90.00</b>  | <b>92.00</b>  | <b>173.00</b> | <b>104.00</b> | <b>746.00</b> |                |
| Exp.         | 59.05         | 112.14        | 53.69         | 54.88         | 103.20        | 62.04         |               | <b>p-value</b> |
| Exp.         | 39.95         | 75.86         | 36.31         | 37.12         | 69.80         | 41.96         |               | 2.70263E-47    |
| X^2          | 1.39          | 8.49          | 44.15         | 25.11         | 7.70          | 5.20          | <b>92.04</b>  |                |
| X^2          | 2.05          | 12.55         | 65.27         | 37.12         | 11.39         | 7.69          | <b>136.08</b> |                |
| CCA          | 81.00         | 47.00         | 90.00         | 152.00        | 93.00         | 221.00        | 684.00        |                |
| CCA          | 7.00          | 0.00          | 34.00         | 108.00        | 16.00         | 42.00         | 207.00        |                |
| <b>TOTAL</b> | <b>88.00</b>  | <b>47.00</b>  | <b>124.00</b> | <b>260.00</b> | <b>109.00</b> | <b>263.00</b> | <b>891.00</b> |                |
| Exp.         | 67.56         | 36.08         | 95.19         | 199.60        | 83.68         | 201.90        |               | <b>p-value</b> |
| Exp.         | 20.44         | 10.92         | 28.81         | 60.40         | 25.32         | 61.10         |               | 1.71467E-17    |
| X^2          | 2.68          | 3.30          | 0.28          | 11.35         | 1.04          | 1.81          | <b>20.46</b>  |                |
| X^2          | 8.84          | 10.92         | 0.94          | 37.50         | 3.43          | 5.97          | <b>67.60</b>  |                |
| CCC          | 6.00          | 150.00        | 0.00          | 211.00        | 1.00          | 93.00         | 461.00        |                |
| CCC          | 0.00          | 5.00          | 0.00          | 91.00         | 0.00          | 42.00         | 138.00        |                |
| <b>TOTAL</b> | <b>6.00</b>   | <b>155.00</b> | <b>0.00</b>   | <b>302.00</b> | <b>1.00</b>   | <b>135.00</b> | <b>599.00</b> |                |
| Exp.         | 4.62          | 119.29        | 0.25          | 232.42        | 0.77          | 103.90        |               | <b>p-value</b> |
| Exp.         | 1.38          | 35.71         | 0.25          | 69.58         | 0.25          | 31.10         |               | 1.58567E-09*   |
| X^2          | 0.41          | 7.91          | 0.25          | 1.97          | 0.07          | 1.14          | <b>11.76</b>  |                |
| X^2          | 1.38          | 26.41         | 0.25          | 6.60          | 0.25          | 3.82          | <b>38.71</b>  |                |
| CCT          | 41.00         | 20.00         | 127.00        | 7.00          | 108.00        | 3.00          | 306.00        |                |
| CCT          | 15.00         | 36.00         | 6.00          | 15.00         | 39.00         | 76.00         | 187.00        |                |
| <b>TOTAL</b> | <b>56.00</b>  | <b>56.00</b>  | <b>133.00</b> | <b>22.00</b>  | <b>147.00</b> | <b>79.00</b>  | <b>493.00</b> |                |
| Exp.         | 34.76         | 34.76         | 82.55         | 13.66         | 91.24         | 49.03         |               | <b>p-value</b> |
| Exp.         | 21.24         | 21.24         | 50.45         | 8.34          | 55.76         | 29.97         |               | 4.30325E-44    |
| X^2          | 1.12          | 6.27          | 23.93         | 3.24          | 3.08          | 43.22         | <b>80.86</b>  |                |
| X^2          | 1.83          | 10.25         | 39.16         | 5.31          | 5.04          | 70.72         | <b>132.32</b> |                |
| CCG          | 40.00         | 1.00          | 6.00          | 1.00          | 21.00         | 0.00          | 69.00         |                |
| CCG          | 3.00          | 32.00         | 16.00         | 0.00          | 2.00          | 32.00         | 85.00         |                |
| <b>TOTAL</b> | <b>43.00</b>  | <b>33.00</b>  | <b>22.00</b>  | <b>1.00</b>   | <b>23.00</b>  | <b>32.00</b>  | <b>154.00</b> |                |
| Exp.         | 19.27         | 14.79         | 9.86          | 0.45          | 10.31         | 14.34         |               | <b>p-value</b> |
| Exp.         | 23.73         | 18.21         | 12.14         | 0.55          | 12.69         | 17.66         |               | 6.55286E-23    |
| X^2          | 22.31         | 12.85         | 1.51          | 0.68          | 11.10         | 14.34         | <b>62.79</b>  |                |
| X^2          | 18.11         | 10.43         | 1.23          | 0.55          | 9.01          | 11.64         | <b>50.97</b>  |                |
| CAT          | 48.00         | 144.00        | 72.00         | 63.00         | 184.00        | 23.00         | 534.00        |                |
| CAT          | 36.00         | 84.00         | 21.00         | 3.00          | 66.00         | 21.00         | 231.00        |                |
| <b>TOTAL</b> | <b>84.00</b>  | <b>228.00</b> | <b>93.00</b>  | <b>66.00</b>  | <b>250.00</b> | <b>44.00</b>  | <b>765.00</b> |                |
| Exp.         | 58.64         | 159.15        | 64.92         | 46.07         | 174.51        | 30.71         |               | <b>p-value</b> |
| Exp.         | 25.36         | 68.85         | 28.08         | 19.93         | 75.49         | 13.29         |               | 4.77169E-08    |
| X^2          | 1.93          | 1.44          | 0.77          | 6.22          | 0.52          | 1.94          | <b>12.82</b>  |                |
| X^2          | 4.46          | 3.34          | 1.79          | 14.38         | 1.19          | 4.48          | <b>29.63</b>  |                |
| CTA          | 82.00         | 52.00         | 1.00          | 27.00         | 1.00          | 2.00          | 165.00        |                |
| CTA          | 20.00         | 2.00          | 40.00         | 16.00         | 45.00         | 9.00          | 132.00        |                |
| <b>TOTAL</b> | <b>102.00</b> | <b>54.00</b>  | <b>41.00</b>  | <b>43.00</b>  | <b>46.00</b>  | <b>11.00</b>  | <b>297.00</b> |                |
| Exp.         | 56.67         | 30.00         | 22.78         | 23.89         | 25.56         | 6.11          |               | <b>p-value</b> |
| Exp.         | 45.33         | 24.00         | 18.22         | 19.11         | 20.44         | 4.89          |               | 1.28144E-34    |
| X^2          | 11.33         | 16.13         | 20.82         | 0.41          | 23.59         | 2.77          | <b>75.05</b>  |                |
| X^2          | 14.16         | 20.17         | 26.03         | 0.51          | 29.49         | 3.46          | <b>93.81</b>  |                |
| CTT          | 98.00         | 8.00          | 4.00          | 3.00          | 58.00         | 43.00         | 214.00        |                |
| CTT          | 56.00         | 47.00         | 24.00         | 0.00          | 35.00         | 32.00         | 194.00        |                |
| <b>TOTAL</b> | <b>154.00</b> | <b>55.00</b>  | <b>28.00</b>  | <b>3.00</b>   | <b>93.00</b>  | <b>75.00</b>  | <b>408.00</b> |                |
| Exp.         | 80.77         | 28.85         | 14.69         | 1.57          | 48.78         | 39.34         |               | <b>p-value</b> |
| Exp.         | 73.23         | 26.15         | 13.31         | 1.43          | 44.22         | 35.66         |               | 3.10181E-12    |
| X^2          | 3.67          | 15.07         | 7.78          | 1.29          | 1.74          | 0.34          | <b>29.89</b>  |                |
| X^2          | 4.05          | 16.62         | 8.58          | 1.43          | 1.92          | 0.38          | <b>32.97</b>  |                |
| CTC          | 0.00          | 62.00         | 4.00          | 174.00        | 0.00          | 15.00         | 255.00        |                |
| CTC          | 0.00          | 12.00         | 1.00          | 24.00         | 0.00          | 24.00         | 61.00         |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>74.00</b>  | <b>5.00</b>   | <b>198.00</b> | <b>0.00</b>   | <b>39.00</b>  | <b>316.00</b> |                |
| Exp.         | 0.25          | 59.72         | 4.03          | 159.78        | 0.25          | 31.47         |               | <b>p-value</b> |

|       |        |        |        |        |        |        |        |              |
|-------|--------|--------|--------|--------|--------|--------|--------|--------------|
| Exp.  | 0.25   | 14.28  | 0.97   | 38.22  | 0.25   | 7.53   |        | 6.30261E-10* |
| X^2   | 0.25   | 0.09   | 0.00   | 1.27   | 0.25   | 8.62   | 10.47  |              |
| X^2   | 0.25   | 0.37   | 0.00   | 5.29   | 0.25   | 36.04  | 42.20  |              |
| CTG   | 144.00 | 0.00   | 2.00   | 0.00   | 0.00   | 0.00   | 146.00 |              |
| CTG   | 115.00 | 0.00   | 45.00  | 0.00   | 28.00  | 4.00   | 192.00 |              |
| TOTAL | 259.00 | 0.00   | 47.00  | 0.00   | 28.00  | 4.00   | 338.00 |              |
| Exp.  | 111.88 | 0.25   | 20.30  | 0.25   | 12.09  | 1.73   |        | p-value      |
| Exp.  | 147.12 | 0.25   | 26.70  | 0.25   | 15.91  | 2.27   |        | 1.23144E-13* |
| X^2   | 9.22   | 0.25   | 16.50  | 0.25   | 12.09  | 1.73   | 40.05  |              |
| X^2   | 7.01   | 0.25   | 12.55  | 0.25   | 9.20   | 1.31   | 30.57  |              |
| CAG   | 0.00   | 0.00   | 1.00   | 60.00  | 27.00  | 113.00 | 201.00 |              |
| CAG   | 2.00   | 2.00   | 0.00   | 35.00  | 8.00   | 139.00 | 186.00 |              |
| TOTAL | 2.00   | 2.00   | 1.00   | 95.00  | 35.00  | 252.00 | 387.00 |              |
| Exp.  | 1.04   | 1.04   | 0.52   | 49.34  | 18.18  | 130.88 |        | p-value      |
| Exp.  | 0.96   | 0.96   | 0.48   | 45.66  | 16.82  | 121.12 |        | 0.000214204* |
| X^2   | 1.04   | 1.04   | 0.44   | 2.30   | 4.28   | 2.44   | 11.55  |              |
| X^2   | 1.12   | 1.12   | 0.48   | 2.49   | 4.63   | 2.64   | 12.48  |              |
| CGA   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   | 0.00   |              |
| CGA   | 3.00   | 0.00   | 61.00  | 0.00   | 25.00  | 0.00   | 89.00  |              |
| TOTAL | 3.00   | 0.00   | 61.00  | 0.00   | 25.00  | 0.00   | 89.00  |              |
| Exp.  | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    |        | p-value      |
| Exp.  | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    |        | N/A          |
| X^2   | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    |              |
| X^2   | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    | N/A    |              |
| CGG   | 1.00   | 0.00   | 38.00  | 0.00   | 2.00   | 0.00   | 41.00  |              |
| CGG   | 11.00  | 2.00   | 1.00   | 0.00   | 31.00  | 0.00   | 45.00  |              |
| TOTAL | 12.00  | 2.00   | 39.00  | 0.00   | 33.00  | 0.00   | 86.00  |              |
| Exp.  | 5.72   | 0.95   | 18.59  | 0.25   | 15.73  | 0.25   |        | p-value      |
| Exp.  | 6.28   | 1.05   | 20.41  | 0.25   | 17.27  | 0.25   |        | 6.69545E-14* |
| X^2   | 3.90   | 0.95   | 20.26  | 0.25   | 11.99  | 0.25   | 37.59  |              |
| X^2   | 3.55   | 0.87   | 18.46  | 0.25   | 10.92  | 0.25   | 34.30  |              |
| CGC   | 10.00  | 1.00   | 1.00   | 42.00  | 0.00   | 1.00   | 55.00  |              |
| CGC   | 1.00   | 1.00   | 53.00  | 52.00  | 32.00  | 33.00  | 172.00 |              |
| TOTAL | 11.00  | 2.00   | 54.00  | 94.00  | 32.00  | 34.00  | 227.00 |              |
| Exp.  | 2.67   | 0.48   | 13.08  | 22.78  | 7.75   | 8.24   |        | p-value      |
| Exp.  | 8.33   | 1.52   | 40.92  | 71.22  | 24.25  | 25.76  |        | 2.99927E-16  |
| X^2   | 20.19  | 0.55   | 11.16  | 16.23  | 7.75   | 6.36   | 62.23  |              |
| X^2   | 6.45   | 0.18   | 3.57   | 5.19   | 2.48   | 2.03   | 19.90  |              |
| CGT   | 47.00  | 5.00   | 5.00   | 30.00  | 0.00   | 2.00   | 89.00  |              |
| CGT   | 0.00   | 0.00   | 21.00  | 0.00   | 21.00  | 0.00   | 42.00  |              |
| TOTAL | 47.00  | 5.00   | 26.00  | 30.00  | 21.00  | 2.00   | 131.00 |              |
| Exp.  | 31.93  | 3.40   | 17.66  | 20.38  | 14.27  | 1.36   |        | p-value      |
| Exp.  | 15.07  | 1.60   | 8.34   | 9.62   | 6.73   | 0.64   |        | 1.23753E-22* |
| X^2   | 7.11   | 0.76   | 9.08   | 4.54   | 14.27  | 0.30   | 36.06  |              |
| X^2   | 15.07  | 1.60   | 19.24  | 9.62   | 30.23  | 0.64   | 76.40  |              |
| TAA   | 1.00   | 12.00  | 0.00   | 102.00 | 3.00   | 36.00  | 154.00 |              |
| TAA   | 35.00  | 110.00 | 5.00   | 121.00 | 0.00   | 37.00  | 308.00 |              |
| TOTAL | 36.00  | 122.00 | 5.00   | 223.00 | 3.00   | 73.00  | 462.00 |              |
| Exp.  | 12.00  | 40.67  | 1.67   | 74.33  | 1.00   | 24.33  |        | p-value      |
| Exp.  | 24.00  | 81.33  | 3.33   | 148.67 | 2.00   | 48.67  |        | 2.45186E-15* |
| X^2   | 10.08  | 20.21  | 1.67   | 10.30  | 4.00   | 5.59   | 51.85  |              |
| X^2   | 5.04   | 10.10  | 0.83   | 5.15   | 2.00   | 2.80   | 25.92  |              |
| TAC   | 335.00 | 59.00  | 130.00 | 37.00  | 9.00   | 8.00   | 578.00 |              |
| TAC   | 70.00  | 2.00   | 58.00  | 1.00   | 3.00   | 1.00   | 135.00 |              |
| TOTAL | 405.00 | 61.00  | 188.00 | 38.00  | 12.00  | 9.00   | 713.00 |              |
| Exp.  | 328.32 | 49.45  | 152.40 | 30.81  | 9.73   | 7.30   |        | p-value      |
| Exp.  | 76.68  | 11.55  | 35.60  | 7.19   | 2.27   | 1.70   |        | 1.45071E-06  |
| X^2   | 0.14   | 1.84   | 3.29   | 1.25   | 0.05   | 0.07   | 6.64   |              |
| X^2   | 0.58   | 7.90   | 14.10  | 5.33   | 0.23   | 0.29   | 28.44  |              |
| TCA   | 1.00   | 3.00   | 19.00  | 26.00  | 320.00 | 7.00   | 376.00 |              |
| TCA   | 14.00  | 0.00   | 67.00  | 0.00   | 34.00  | 0.00   | 115.00 |              |
| TOTAL | 15.00  | 3.00   | 86.00  | 26.00  | 354.00 | 7.00   | 491.00 |              |
| Exp.  | 11.49  | 2.30   | 65.86  | 19.91  | 271.09 | 5.36   |        | p-value      |
| Exp.  | 3.51   | 0.70   | 20.14  | 6.09   | 82.91  | 1.64   |        | 4.16722E-48* |
| X^2   | 9.57   | 0.21   | 33.34  | 1.86   | 8.83   | 0.50   | 54.32  |              |
| X^2   | 31.30  | 0.70   | 109.00 | 6.09   | 28.85  | 1.64   | 177.59 |              |
| TCC   | 6.00   | 29.00  | 73.00  | 16.00  | 73.00  | 5.00   | 202.00 |              |

Characterizing patterns in DNA sequence trace data through informatics tools

|       |        |        |        |        |        |        |        |              |
|-------|--------|--------|--------|--------|--------|--------|--------|--------------|
| TCC   | 3.00   | 27.00  | 18.00  | 21.00  | 56.00  | 14.00  | 139.00 |              |
| TOTAL | 9.00   | 56.00  | 91.00  | 37.00  | 129.00 | 19.00  | 341.00 |              |
| Exp.  | 5.33   | 33.17  | 53.91  | 21.92  | 76.42  | 11.26  |        | p-value      |
| Exp.  | 3.67   | 22.83  | 37.09  | 15.08  | 52.58  | 7.74   |        | 9.76724E-06  |
| X^2   | 0.08   | 0.52   | 6.76   | 1.60   | 0.15   | 3.48   | 12.60  |              |
| X^2   | 0.12   | 0.76   | 9.83   | 2.32   | 0.22   | 5.05   | 18.31  |              |
| TCT   | 1.00   | 0.00   | 68.00  | 0.00   | 36.00  | 0.00   | 105.00 |              |
| TCT   | 17.00  | 23.00  | 57.00  | 4.00   | 79.00  | 0.00   | 180.00 |              |
| TOTAL | 18.00  | 23.00  | 125.00 | 4.00   | 115.00 | 0.00   | 285.00 |              |
| Exp.  | 6.63   | 8.47   | 46.05  | 1.47   | 42.37  | 0.25   |        | p-value      |
| Exp.  | 11.37  | 14.53  | 78.95  | 2.53   | 72.63  | 0.25   |        | 7.79446E-08* |
| X^2   | 4.78   | 8.47   | 10.46  | 1.47   | 0.96   | 0.25   | 26.40  |              |
| X^2   | 2.79   | 4.94   | 6.10   | 0.86   | 0.56   | 0.25   | 15.50  |              |
| TCG   | 1.00   | 0.00   | 1.00   | 0.00   | 67.00  | 6.00   | 75.00  |              |
| TCG   | 0.00   | 0.00   | 2.00   | 1.00   | 33.00  | 42.00  | 78.00  |              |
| TOTAL | 1.00   | 0.00   | 3.00   | 1.00   | 100.00 | 48.00  | 153.00 |              |
| Exp.  | 0.49   | 0.25   | 1.47   | 0.49   | 49.02  | 23.53  |        | p-value      |
| Exp.  | 0.51   | 0.25   | 1.53   | 0.51   | 50.98  | 24.47  |        | 1.00597E-07* |
| X^2   | 0.53   | 0.25   | 0.15   | 0.49   | 6.60   | 13.06  | 21.08  |              |
| X^2   | 0.51   | 0.25   | 0.14   | 0.47   | 6.34   | 12.56  | 20.27  |              |
| TAT   | 36.00  | 35.00  | 110.00 | 29.00  | 17.00  | 1.00   | 228.00 |              |
| TAT   | 73.00  | 102.00 | 10.00  | 16.00  | 2.00   | 8.00   | 211.00 |              |
| TOTAL | 109.00 | 137.00 | 120.00 | 45.00  | 19.00  | 9.00   | 439.00 |              |
| Exp.  | 56.61  | 71.15  | 62.32  | 23.37  | 9.87   | 4.67   |        | p-value      |
| Exp.  | 52.39  | 65.85  | 57.68  | 21.63  | 9.13   | 4.33   |        | 1.91224E-30  |
| X^2   | 7.50   | 18.37  | 36.47  | 1.36   | 5.15   | 2.89   | 71.74  |              |
| X^2   | 8.11   | 19.85  | 39.41  | 1.46   | 5.57   | 3.12   | 77.52  |              |
| TTA   | 21.00  | 9.00   | 108.00 | 5.00   | 22.00  | 26.00  | 191.00 |              |
| TTA   | 104.00 | 26.00  | 113.00 | 5.00   | 9.00   | 13.00  | 270.00 |              |
| TOTAL | 125.00 | 35.00  | 221.00 | 10.00  | 31.00  | 39.00  | 461.00 |              |
| Exp.  | 51.79  | 14.50  | 91.56  | 4.14   | 12.84  | 16.16  |        | p-value      |
| Exp.  | 73.21  | 20.50  | 129.44 | 5.86   | 18.16  | 22.84  |        | 5.84873E-12  |
| X^2   | 18.30  | 2.09   | 2.95   | 0.18   | 6.53   | 5.99   | 36.04  |              |
| X^2   | 12.95  | 1.48   | 2.09   | 0.13   | 4.62   | 4.24   | 25.50  |              |
| TTT   | 20.00  | 11.00  | 1.00   | 24.00  | 7.00   | 76.00  | 139.00 |              |
| TTT   | 1.00   | 41.00  | 0.00   | 81.00  | 0.00   | 71.00  | 194.00 |              |
| TOTAL | 21.00  | 52.00  | 1.00   | 105.00 | 7.00   | 147.00 | 333.00 |              |
| Exp.  | 8.77   | 21.71  | 0.42   | 43.83  | 2.92   | 61.36  |        | p-value      |
| Exp.  | 12.23  | 30.29  | 0.58   | 61.17  | 4.08   | 85.64  |        | 5.91723E-13* |
| X^2   | 14.40  | 5.28   | 0.81   | 8.97   | 5.69   | 3.49   | 38.65  |              |
| X^2   | 10.32  | 3.78   | 0.58   | 6.43   | 4.08   | 2.50   | 27.69  |              |
| TTC   | 1.00   | 24.00  | 0.00   | 67.00  | 0.00   | 56.00  | 148.00 |              |
| TTC   | 49.00  | 0.00   | 1.00   | 12.00  | 14.00  | 115.00 | 191.00 |              |
| TOTAL | 50.00  | 24.00  | 1.00   | 79.00  | 14.00  | 171.00 | 339.00 |              |
| Exp.  | 21.83  | 10.48  | 0.44   | 34.49  | 6.11   | 74.65  |        | p-value      |
| Exp.  | 28.17  | 13.52  | 0.56   | 44.51  | 7.89   | 96.35  |        | 1.37724E-28  |
| X^2   | 19.87  | 17.45  | 0.44   | 30.64  | 6.11   | 4.66   | 79.18  |              |
| X^2   | 15.40  | 13.52  | 0.34   | 23.75  | 4.74   | 3.61   | 61.35  |              |
| TTG   | 67.00  | 0.00   | 114.00 | 0.00   | 0.00   | 0.00   | 181.00 |              |
| TTG   | 62.00  | 0.00   | 87.00  | 0.00   | 0.00   | 0.00   | 149.00 |              |
| TOTAL | 129.00 | 0.00   | 201.00 | 0.00   | 0.00   | 0.00   | 330.00 |              |
| Exp.  | 70.75  | 0.25   | 110.25 | 0.25   | 0.25   | 0.25   |        | p-value      |
| Exp.  | 58.25  | 0.25   | 90.75  | 0.25   | 0.25   | 0.25   |        | 0.981600137* |
| X^2   | 0.20   | 0.25   | 0.13   | 0.25   | 0.25   | 0.25   | 1.33   |              |
| X^2   | 0.24   | 0.25   | 0.16   | 0.25   | 0.25   | 0.25   | 1.40   |              |
| TAG   | 0.00   | 0.00   | 0.00   | 92.00  | 14.00  | 58.00  | 164.00 |              |
| TAG   | 0.00   | 0.00   | 1.00   | 49.00  | 2.00   | 3.00   | 55.00  |              |
| TOTAL | 0.00   | 0.00   | 1.00   | 141.00 | 16.00  | 61.00  | 219.00 |              |
| Exp.  | 0.25   | 0.25   | 0.75   | 105.59 | 11.98  | 45.68  |        | p-value      |
| Exp.  | 0.25   | 0.25   | 0.25   | 35.41  | 4.02   | 15.32  |        | 0.000171764* |
| X^2   | 0.25   | 0.25   | 0.75   | 1.75   | 0.34   | 3.32   | 6.66   |              |
| X^2   | 0.25   | 0.25   | 2.23   | 5.21   | 1.01   | 9.91   | 18.87  |              |
| TGA   | 28.00  | 74.00  | 0.00   | 67.00  | 0.00   | 0.00   | 169.00 |              |
| TGA   | 0.00   | 74.00  | 0.00   | 7.00   | 0.00   | 0.00   | 81.00  |              |
| TOTAL | 28.00  | 148.00 | 0.00   | 74.00  | 0.00   | 0.00   | 250.00 |              |
| Exp.  | 18.93  | 100.05 | 0.25   | 50.02  | 0.25   | 0.25   |        | p-value      |
| Exp.  | 9.07   | 47.95  | 0.25   | 23.98  | 0.25   | 0.25   |        | 5.0687E-10*  |
| X^2   | 4.35   | 6.78   | 0.25   | 5.76   | 0.25   | 0.25   | 17.64  |              |

|              |               |               |              |               |               |              |               |                |
|--------------|---------------|---------------|--------------|---------------|---------------|--------------|---------------|----------------|
| X^2          | 9.07          | 14.15         | 0.25         | 12.02         | 0.25          | 0.25         | 35.99         |                |
| TGG          | 0.00          | 59.00         | 0.00         | 29.00         | 0.00          | 0.00         | 88.00         |                |
| TGG          | 0.00          | 33.00         | 0.00         | 44.00         | 0.00          | 4.00         | 81.00         |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>92.00</b>  | <b>0.00</b>  | <b>73.00</b>  | <b>0.00</b>   | <b>4.00</b>  | <b>169.00</b> |                |
| Exp.         | 0.25          | 47.91         | 0.25         | 38.01         | 0.25          | 2.08         |               | <b>p-value</b> |
| Exp.         | 0.25          | 44.09         | 0.25         | 34.99         | 0.25          | 1.92         |               | 0.014598301*   |
| X^2          | 0.25          | 2.57          | 0.25         | 2.14          | 0.25          | 2.08         | 7.54          |                |
| X^2          | 0.25          | 2.79          | 0.25         | 2.32          | 0.25          | 2.26         | 8.13          |                |
| TGC          | 4.00          | 34.00         | 0.00         | 85.00         | 0.00          | 0.00         | 123.00        |                |
| TGC          | 22.00         | 65.00         | 1.00         | 91.00         | 0.00          | 0.00         | 179.00        |                |
| <b>TOTAL</b> | <b>26.00</b>  | <b>99.00</b>  | <b>1.00</b>  | <b>176.00</b> | <b>0.00</b>   | <b>0.00</b>  | <b>302.00</b> |                |
| Exp.         | 10.59         | 40.32         | 0.41         | 71.68         | 0.25          | 0.25         |               | <b>p-value</b> |
| Exp.         | 15.41         | 58.68         | 0.59         | 104.32        | 0.25          | 0.25         |               | 0.019495242*   |
| X^2          | 4.10          | 0.99          | 0.41         | 2.47          | 0.25          | 0.25         | 8.47          |                |
| X^2          | 2.82          | 0.68          | 0.28         | 1.70          | 0.25          | 0.25         | 5.98          |                |
| TGT          | 0.00          | 1.00          | 0.00         | 135.00        | 0.00          | 1.00         | 137.00        |                |
| TGT          | 0.00          | 21.00         | 0.00         | 200.00        | 0.00          | 0.00         | 221.00        |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>22.00</b>  | <b>0.00</b>  | <b>335.00</b> | <b>0.00</b>   | <b>1.00</b>  | <b>358.00</b> |                |
| Exp.         | 0.25          | 8.42          | 0.25         | 128.20        | 0.25          | 0.38         |               | <b>p-value</b> |
| Exp.         | 0.25          | 13.58         | 0.25         | 206.80        | 0.25          | 0.62         |               | 0.025445335*   |
| X^2          | 0.25          | 6.54          | 0.25         | 0.36          | 0.25          | 1.00         | 8.64          |                |
| X^2          | 0.25          | 4.05          | 0.25         | 0.22          | 0.25          | 0.62         | 5.64          |                |
| GAA          | 35.00         | 0.00          | 0.00         | 0.00          | 29.00         | 10.00        | 74.00         |                |
| GAA          | 1.00          | 43.00         | 0.00         | 2.00          | 5.00          | 34.00        | 85.00         |                |
| <b>TOTAL</b> | <b>36.00</b>  | <b>43.00</b>  | <b>0.00</b>  | <b>2.00</b>   | <b>34.00</b>  | <b>44.00</b> | <b>159.00</b> |                |
| Exp.         | 16.75         | 20.01         | 0.25         | 0.93          | 15.82         | 20.48        |               | <b>p-value</b> |
| Exp.         | 19.25         | 22.99         | 0.25         | 1.07          | 18.18         | 23.52        |               | 1.85645E-21*   |
| X^2          | 19.87         | 20.01         | 0.25         | 0.93          | 10.97         | 5.36         | 57.39         |                |
| X^2          | 17.30         | 17.42         | 0.25         | 0.81          | 9.55          | 4.67         | 50.00         |                |
| GAC          | 0.00          | 1.00          | 0.00         | 26.00         | 57.00         | 41.00        | 125.00        |                |
| GAC          | 11.00         | 57.00         | 0.00         | 28.00         | 0.00          | 1.00         | 97.00         |                |
| <b>TOTAL</b> | <b>11.00</b>  | <b>58.00</b>  | <b>0.00</b>  | <b>54.00</b>  | <b>57.00</b>  | <b>42.00</b> | <b>222.00</b> |                |
| Exp.         | 6.19          | 32.66         | 0.25         | 30.41         | 32.09         | 23.65        |               | <b>p-value</b> |
| Exp.         | 4.81          | 25.34         | 0.25         | 23.59         | 24.91         | 18.35        |               | 1.4373E-32     |
| X^2          | 6.19          | 30.69         | 0.25         | 0.64          | 19.33         | 12.73        | 69.83         |                |
| X^2          | 7.98          | 39.55         | 0.25         | 0.82          | 24.91         | 16.41        | 89.91         |                |
| GCA          | 33.00         | 37.00         | 13.00        | 36.00         | 33.00         | 36.00        | 188.00        |                |
| GCA          | 38.00         | 65.00         | 46.00        | 23.00         | 1.00          | 45.00        | 218.00        |                |
| <b>TOTAL</b> | <b>71.00</b>  | <b>102.00</b> | <b>59.00</b> | <b>59.00</b>  | <b>34.00</b>  | <b>81.00</b> | <b>406.00</b> |                |
| Exp.         | 32.88         | 47.23         | 27.32        | 27.32         | 15.74         | 37.51        |               | <b>p-value</b> |
| Exp.         | 38.12         | 54.77         | 31.68        | 31.68         | 18.26         | 43.49        |               | 2.3865E-11     |
| X^2          | 0.00          | 2.22          | 7.51         | 2.76          | 18.91         | 0.06         | 31.45         |                |
| X^2          | 0.00          | 1.91          | 6.47         | 2.38          | 16.31         | 0.05         | 27.13         |                |
| GCC          | 97.00         | 16.00         | 3.00         | 0.00          | 41.00         | 2.00         | 159.00        |                |
| GCC          | 18.00         | 32.00         | 39.00        | 0.00          | 64.00         | 1.00         | 154.00        |                |
| <b>TOTAL</b> | <b>115.00</b> | <b>48.00</b>  | <b>42.00</b> | <b>0.00</b>   | <b>105.00</b> | <b>3.00</b>  | <b>313.00</b> |                |
| Exp.         | 58.42         | 24.38         | 21.34        | 0.25          | 53.34         | 1.52         |               | <b>p-value</b> |
| Exp.         | 56.58         | 23.62         | 20.66        | 0.25          | 51.66         | 1.48         |               | 4.09952E-19*   |
| X^2          | 25.48         | 2.88          | 15.76        | 0.25          | 2.85          | 0.15         | 47.37         |                |
| X^2          | 26.31         | 2.98          | 16.27        | 0.25          | 2.95          | 0.15         | 48.90         |                |
| GCT          | 0.00          | 0.00          | 1.00         | 4.00          | 40.00         | 34.00        | 79.00         |                |
| GCT          | 0.00          | 31.00         | 51.00        | 0.00          | 1.00          | 32.00        | 115.00        |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>31.00</b>  | <b>52.00</b> | <b>4.00</b>   | <b>41.00</b>  | <b>66.00</b> | <b>194.00</b> |                |
| Exp.         | 0.25          | 12.62         | 21.18        | 1.63          | 16.70         | 26.88        |               | <b>p-value</b> |
| Exp.         | 0.25          | 18.38         | 30.82        | 2.37          | 24.30         | 39.12        |               | 1.00929E-23*   |
| X^2          | 0.25          | 12.62         | 19.22        | 3.45          | 32.53         | 1.89         | 69.96         |                |
| X^2          | 0.25          | 8.67          | 13.21        | 2.37          | 22.35         | 1.30         | 48.14         |                |
| GCG          | 0.00          | 0.00          | 0.00         | 0.00          | 0.00          | 0.00         | 0.00          |                |
| GCG          | 0.00          | 2.00          | 13.00        | 23.00         | 1.00          | 57.00        | 96.00         |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>2.00</b>   | <b>13.00</b> | <b>23.00</b>  | <b>1.00</b>   | <b>57.00</b> | <b>96.00</b>  |                |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A          |               | <b>p-value</b> |
| Exp.         | N/A           | N/A           | N/A          | N/A           | N/A           | N/A          |               | N/A            |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A          | N/A           |                |
| X^2          | N/A           | N/A           | N/A          | N/A           | N/A           | N/A          | N/A           |                |
| GAT          | 0.00          | 63.00         | 0.00         | 54.00         | 0.00          | 6.00         | 123.00        |                |
| GAT          | 0.00          | 42.00         | 0.00         | 0.00          | 0.00          | 42.00        | 84.00         |                |
| <b>TOTAL</b> | <b>0.00</b>   | <b>105.00</b> | <b>0.00</b>  | <b>54.00</b>  | <b>0.00</b>   | <b>48.00</b> | <b>207.00</b> |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|                |               |              |              |              |               |              |               |                |
|----------------|---------------|--------------|--------------|--------------|---------------|--------------|---------------|----------------|
| Exp.           | 0.25          | 62.39        | 0.25         | 32.09        | 0.25          | 28.52        |               | <b>p-value</b> |
| Exp.           | 0.25          | 42.61        | 0.25         | 21.91        | 0.25          | 19.48        |               | 5.93954E-16*   |
| X <sup>2</sup> | 0.25          | 0.01         | 0.25         | 14.97        | 0.25          | 17.78        | <b>33.50</b>  |                |
| X <sup>2</sup> | 0.25          | 0.01         | 0.25         | 21.91        | 0.25          | 26.04        | <b>48.71</b>  |                |
| GTA            | 124.00        | 43.00        | 24.00        | 61.00        | 98.00         | 92.00        | 442.00        |                |
| GTA            | 25.00         | 9.00         | 2.00         | 0.00         | 41.00         | 6.00         | 83.00         |                |
| <b>TOTAL</b>   | <b>149.00</b> | <b>52.00</b> | <b>26.00</b> | <b>61.00</b> | <b>139.00</b> | <b>98.00</b> | <b>525.00</b> |                |
| Exp.           | 125.44        | 43.78        | 21.89        | 51.36        | 117.02        | 82.51        |               | <b>p-value</b> |
| Exp.           | 23.56         | 8.22         | 4.11         | 9.64         | 21.98         | 15.49        |               | 1.96631E-07    |
| X <sup>2</sup> | 0.02          | 0.01         | 0.20         | 1.81         | 3.09          | 1.09         | <b>6.23</b>   |                |
| X <sup>2</sup> | 0.09          | 0.07         | 1.08         | 9.64         | 16.47         | 5.82         | <b>33.18</b>  |                |
| GTT            | 0.00          | 0.00         | 0.00         | 0.00         | 22.00         | 0.00         | 22.00         |                |
| GTT            | 0.00          | 28.00        | 0.00         | 5.00         | 2.00          | 32.00        | 67.00         |                |
| <b>TOTAL</b>   | <b>0.00</b>   | <b>28.00</b> | <b>0.00</b>  | <b>5.00</b>  | <b>24.00</b>  | <b>32.00</b> | <b>89.00</b>  |                |
| Exp.           | 0.25          | 6.92         | 0.25         | 1.24         | 5.93          | 7.91         |               | <b>p-value</b> |
| Exp.           | 0.25          | 21.08        | 0.25         | 3.76         | 18.07         | 24.09        |               | 1.26492E-15*   |
| X <sup>2</sup> | 0.25          | 6.92         | 0.25         | 1.24         | 43.52         | 7.91         | <b>60.08</b>  |                |
| X <sup>2</sup> | 0.25          | 2.27         | 0.25         | 0.41         | 14.29         | 2.60         | <b>20.06</b>  |                |
| GTC            | 22.00         | 10.00        | 9.00         | 4.00         | 3.00          | 24.00        | 72.00         |                |
| GTC            | 10.00         | 0.00         | 0.00         | 0.00         | 34.00         | 66.00        | 110.00        |                |
| <b>TOTAL</b>   | <b>32.00</b>  | <b>10.00</b> | <b>9.00</b>  | <b>4.00</b>  | <b>37.00</b>  | <b>90.00</b> | <b>182.00</b> |                |
| Exp.           | 12.66         | 3.96         | 3.56         | 1.58         | 14.64         | 35.60        |               | <b>p-value</b> |
| Exp.           | 19.34         | 6.04         | 5.44         | 2.42         | 22.36         | 54.40        |               | 2.53594E-13*   |
| X <sup>2</sup> | 6.89          | 9.23         | 8.31         | 3.69         | 9.25          | 3.78         | <b>41.16</b>  |                |
| X <sup>2</sup> | 4.51          | 6.04         | 5.44         | 2.42         | 6.06          | 2.48         | <b>26.94</b>  |                |
| GTG            | 0.00          | 0.00         | 0.00         | 0.00         | 0.00          | 0.00         | 0.00          |                |
| GTG            | 10.00         | 0.00         | 44.00        | 0.00         | 54.00         | 0.00         | 108.00        |                |
| <b>TOTAL</b>   | <b>10.00</b>  | <b>0.00</b>  | <b>44.00</b> | <b>0.00</b>  | <b>54.00</b>  | <b>0.00</b>  | <b>108.00</b> |                |
| Exp.           | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          |               | <b>p-value</b> |
| Exp.           | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          |               | N/A            |
| X <sup>2</sup> | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          | <b>N/A</b>    |                |
| X <sup>2</sup> | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          | <b>N/A</b>    |                |
| GAG            | 0.00          | 0.00         | 0.00         | 0.00         | 0.00          | 0.00         | 0.00          |                |
| GAG            | 8.00          | 20.00        | 0.00         | 36.00        | 0.00          | 36.00        | 100.00        |                |
| <b>TOTAL</b>   | <b>8.00</b>   | <b>20.00</b> | <b>0.00</b>  | <b>36.00</b> | <b>0.00</b>   | <b>36.00</b> | <b>100.00</b> |                |
| Exp.           | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          |               | <b>p-value</b> |
| Exp.           | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          |               | N/A            |
| X <sup>2</sup> | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          | <b>N/A</b>    |                |
| X <sup>2</sup> | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          | <b>N/A</b>    |                |
| GGA            | 66.00         | 1.00         | 0.00         | 0.00         | 22.00         | 5.00         | 94.00         |                |
| GGA            | 37.00         | 22.00        | 0.00         | 2.00         | 19.00         | 13.00        | 93.00         |                |
| <b>TOTAL</b>   | <b>103.00</b> | <b>23.00</b> | <b>0.00</b>  | <b>2.00</b>  | <b>41.00</b>  | <b>18.00</b> | <b>187.00</b> |                |
| Exp.           | 51.78         | 11.56        | 0.25         | 1.01         | 20.61         | 9.05         |               | <b>p-value</b> |
| Exp.           | 51.22         | 11.44        | 0.25         | 0.99         | 20.39         | 8.95         |               | 3.57942E-06*   |
| X <sup>2</sup> | 3.91          | 9.65         | 0.25         | 1.01         | 0.09          | 1.81         | <b>16.72</b>  |                |
| X <sup>2</sup> | 3.95          | 9.75         | 0.25         | 1.02         | 0.09          | 1.83         | <b>16.89</b>  |                |
| GGG            | 1.00          | 30.00        | 0.00         | 0.00         | 0.00          | 62.00        | 93.00         |                |
| GGG            | 1.00          | 1.00         | 0.00         | 0.00         | 0.00          | 18.00        | 20.00         |                |
| <b>TOTAL</b>   | <b>2.00</b>   | <b>31.00</b> | <b>0.00</b>  | <b>0.00</b>  | <b>0.00</b>   | <b>80.00</b> | <b>113.00</b> |                |
| Exp.           | 1.65          | 25.51        | 0.25         | 0.25         | 0.25          | 65.84        |               | <b>p-value</b> |
| Exp.           | 0.35          | 5.49         | 0.25         | 0.25         | 0.25          | 14.16        |               | 0.209268032*   |
| X <sup>2</sup> | 0.25          | 0.79         | 0.25         | 0.25         | 0.25          | 0.22         | <b>2.02</b>   |                |
| X <sup>2</sup> | 1.18          | 3.67         | 0.25         | 0.25         | 0.25          | 1.04         | <b>6.64</b>   |                |
| GGC            | 0.00          | 0.00         | 0.00         | 0.00         | 0.00          | 0.00         | 0.00          |                |
| GGC            | 0.00          | 29.00        | 0.00         | 1.00         | 0.00          | 32.00        | 62.00         |                |
| <b>TOTAL</b>   | <b>0.00</b>   | <b>29.00</b> | <b>0.00</b>  | <b>1.00</b>  | <b>0.00</b>   | <b>32.00</b> | <b>62.00</b>  |                |
| Exp.           | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          |               | <b>p-value</b> |
| Exp.           | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          |               | N/A            |
| X <sup>2</sup> | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          | <b>N/A</b>    |                |
| X <sup>2</sup> | N/A           | N/A          | N/A          | N/A          | N/A           | N/A          | <b>N/A</b>    |                |
| GGT            | 0.00          | 0.00         | 2.00         | 41.00        | 21.00         | 39.00        | 103.00        |                |
| GGT            | 0.00          | 0.00         | 0.00         | 33.00        | 0.00          | 2.00         | 35.00         |                |
| <b>TOTAL</b>   | <b>0.00</b>   | <b>0.00</b>  | <b>2.00</b>  | <b>74.00</b> | <b>21.00</b>  | <b>41.00</b> | <b>138.00</b> |                |
| Exp.           | 0.25          | 0.25         | 1.49         | 55.23        | 15.67         | 30.60        |               | <b>p-value</b> |
| Exp.           | 0.25          | 0.25         | 0.51         | 18.77        | 5.33          | 10.40        |               | 7.94153E-06*   |
| X <sup>2</sup> | 0.25          | 0.25         | 0.17         | 3.67         | 1.81          | 2.30         | <b>8.45</b>   |                |
| X <sup>2</sup> | 0.25          | 0.25         | 0.51         | 10.79        | 5.33          | 6.78         | <b>23.91</b>  |                |

\*Kolmogorov-Smirnov test performed and generally consistent with the Chi Square test

**Table B.** The p-values for the comparisons between Primers A1 and C1; Primers B1 and D1; and Primers A1 and B1. Any p-value < 7.8125E-4 (0.05/64) is considered statistically significant with an overall  $\alpha = 0.05$  for that frame given the null hypothesis that the distribution of patterns is the same. All p-values exceeding 7.8125E-4 are black reversed; the null hypothesis is not rejected. Overall, the primer comparisons yielded statistically significant different distributions of patterns. All N/A correspond to no observations or too few observations for that frame.

| Frame | p-value      |              |              |
|-------|--------------|--------------|--------------|
|       | A1 - C1      | B1 - D1      | A1 - B1      |
| AAA   | 1.20022E-07  | 7.013E-09*   | 6.45651E-12  |
| AAC   | 1.40989E-28  | 4.53575E-05* | 1.99847E-15* |
| ACA   | 2.00702E-29  | 3.12266E-06* | 4.73195E-36  |
| ACC   | 7.72392E-15  | 0.934557717* | 5.09304E-33  |
| ACT   | 9.3124E-27   | 7.84194E-25* | 2.11203E-22  |
| ACG   | 3.4923E-25*  | 1.03049E-08* | 3.9667E-08*  |
| AAT   | 9.02911E-26  | 7.98438E-12  | 1.32109E-29  |
| ATA   | 6.74572E-13  | 5.27502E-28  | 1.39138E-15  |
| ATT   | 1.99252E-55  | 9.96064E-11  | 7.82831E-10  |
| ATC   | 2.65293E-13  | 0.004300712* | 1.45441E-07  |
| ATG   | 0.004046262* | 0.002898016  | 2.08447E-25  |
| AAG   | 2.91673E-12  | 2.26737E-17  | 2.85657E-14  |
| AGA   | 0.216618383* | 2.67454E-10* | 2.01334E-11  |
| AGG   | 3.61857E-10* | 1.79521E-06* | 2.97852E-14  |
| AGC   | 2.24005E-09* | 6.06573E-12* | 4.60009E-07* |
| AGT   | 4.9304E-13   | 1.81423E-10* | 1.29992E-21  |
| CAA   | 1.00751E-31  | 1.71861E-07* | 1.95454E-43  |
| CAC   | 2.70263E-47  | N/A          | N/A          |
| CCA   | 1.71467E-17  | 2.59221E-14  | 3.9681E-10   |
| CCC   | 1.58567E-09* | 1.25396E-06* | 5.11253E-06* |
| CCT   | 4.30325E-44  | 3.14998E-07* | 8.43966E-25  |
| CCG   | 6.55286E-23  | 0.000342836* | 9.20901E-19* |
| CAT   | 4.77169E-08  | 5.12204E-11* | 1.49321E-36  |
| CTA   | 1.28144E-34  | 1.59877E-15* | 1.77731E-36  |
| CTT   | 3.10181E-12  | 1.56409E-11* | 8.90771E-45  |
| CTC   | 6.30261E-10* | N/A#         | N/A#         |
| CTG   | 1.23144E-13* | 1.39157E-07* | 3.18246E-19* |
| CAG   | 0.000214204* | 1.02616E-05* | 0.000668257* |
| CGA   | N/A          | 0.160445616* | N/A          |
| CGG   | 6.69545E-14* | 1.30627E-07* | 9.33045E-13* |
| CGC   | 2.99927E-16  | N/A          | N/A          |
| CGT   | 1.23753E-22* | 1.32679E-05* | 2.95144E-21* |
| TAA   | 2.45186E-15* | 1.54798E-28  | 2.62831E-60  |
| TAC   | 1.45071E-06  | 3.5057E-16*  | 3.97032E-21  |
| TCA   | 4.16722E-48* | 6.27949E-10* | 9.52412E-45  |
| TCC   | 9.76724E-06  | 3.30645E-10* | 2.96373E-15  |
| TCT   | 7.79446E-08* | 0.574033387* | 0.083747644* |
| TCG   | 1.00597E-07* | N/A          | N/A          |
| TAT   | 1.91224E-30  | 1.51128E-08* | 5.68855E-11  |
| TTA   | 5.84873E-12  | 3.7283E-15   | 4.46194E-06  |
| TTT   | 5.91723E-13* | 3.35272E-21  | 6.07473E-11  |
| TTC   | 1.37724E-28  | 0.001968687* | 4.12383E-08* |
| TTG   | 0.981600137* | 4.65451E-15  | 4.29383E-23  |
| TAG   | 0.000171764* | 0.068379023* | 0.0008092*   |
| TGA   | 5.0687E-10*  | 8.24932E-10  | 4.14736E-60  |
| TGG   | 0.014598301  | 1.49799E-09* | 2.83406E-07* |
| TGC   | 0.019495242* | 1.35218E-37  | 4.4806E-27   |
| TGT   | 0.025445335* | 1.87482E-10  | 2.07463E-11  |
| GAA   | 1.85645E-21* | 2.66916E-11* | 2.57306E-30  |
| GAC   | 1.4373E-32   | 1.76059E-24* | 8.6831E-34*  |
| GCA   | 2.3865E-11   | 2.21577E-12  | 5.24235E-06  |
| GCC   | 4.09952E-19* | N/A          | N/A          |
| GCT   | 1.00929E-23* | 6.53521E-16  | 2.35488E-26  |
| GCG   | N/A          | 4.30287E-10* | N/A          |
| GAT   | 5.93954E-16* | 3.21042E-21  | 1.04213E-08  |
| GTA   | 1.96631E-07  | 2.77654E-10  | 1.58E-28     |
| GTT   | 1.26492E-15* | 8.6537E-36   | 2.48183E-09* |
| GTC   | 2.53594E-13* | 2.733E-25*   | 3.57506E-17  |
| GTG   | N/A          | 0.000976928* | N/A          |
| GAG   | N/A          | 4.56534E-33  | N/A          |
| GGA   | 3.57942E-06* | 4.99242E-20  | 7.60976E-13* |
| GGG   | 0.209268032* | 4.38836E-12* | 3.8279E-10*  |
| GGC   | N/A          | 3.09797E-28  | N/A          |
| GGT   | 7.94153E-06* | 1.48212E-28  | 1.90916E-08  |

\*Kolmogorov-Smirnov test performed and generally consistent with the Chi Square test.

## Appendix V — Comparisons of Frames to Different Dye Chemistries

The following table is an example of the processed data using the bioinformatics tools to statistically evaluate each of the 64 frames using a 2 x 6 Chi Square analysis comparing the results from data for the same samples using different dye chemistries produced in the same laboratory. These results demonstrate that the distribution of patterns from the different dye chemistries is not equivalent. Comparisons were made between dRhodamine and BigDye v1.1.

**Table A.** 2 x 6 Chi-square analysis comparing the results between dRhodamine and BigDye v1.1 for Primer C1.

|                |               |               |               |               |              |              |               |                |
|----------------|---------------|---------------|---------------|---------------|--------------|--------------|---------------|----------------|
|                | <b>A</b>      | <b>B</b>      | <b>C</b>      | <b>D</b>      | <b>E</b>     | <b>F</b>     | <b>TOTAL</b>  |                |
| AAA            | 50.00         | 21.00         | 7.00          | 10.00         | 15.00        | 10.00        | 113.00        |                |
| AAA            | 31.00         | 5.00          | 94.00         | 2.00          | 21.00        | 25.00        | 178.00        |                |
| <b>TOTAL</b>   | <b>81.00</b>  | <b>26.00</b>  | <b>101.00</b> | <b>12.00</b>  | <b>36.00</b> | <b>35.00</b> | <b>291.00</b> |                |
| Exp.           | 31.45         | 10.10         | 39.22         | 4.66          | 13.98        | 13.59        |               | <b>p-value</b> |
| Exp.           | 49.55         | 15.90         | 61.78         | 7.34          | 22.02        | 21.41        |               | 2.45507E-18    |
| X <sup>2</sup> | 10.94         | 11.78         | 26.47         | 6.12          | 0.07         | 0.95         | <b>56.32</b>  |                |
| X <sup>2</sup> | 6.94          | 7.48          | 16.80         | 3.89          | 0.05         | 0.60         | <b>35.76</b>  |                |
| AAC            | 14.00         | 7.00          | 9.00          | 8.00          | 12.00        | 6.00         | 56.00         |                |
| AAC            | 116.00        | 1.00          | 26.00         | 0.00          | 87.00        | 4.00         | 234.00        |                |
| <b>TOTAL</b>   | <b>130.00</b> | <b>8.00</b>   | <b>35.00</b>  | <b>8.00</b>   | <b>99.00</b> | <b>10.00</b> | <b>290.00</b> |                |
| Exp.           | 25.10         | 1.54          | 6.76          | 1.54          | 19.12        | 1.93         |               | <b>p-value</b> |
| Exp.           | 104.90        | 6.46          | 28.24         | 6.46          | 79.88        | 8.07         |               | 1.97785E-15    |
| X <sup>2</sup> | 4.91          | 19.26         | 0.74          | 26.97         | 2.65         | 8.57         | <b>63.12</b>  |                |
| X <sup>2</sup> | 1.18          | 4.61          | 0.18          | 6.46          | 0.63         | 2.05         | <b>15.10</b>  |                |
| ACA            | 6.00          | 6.00          | 12.00         | 17.00         | 7.00         | 12.00        | 60.00         |                |
| ACA            | 31.00         | 166.00        | 1.00          | 122.00        | 32.00        | 27.00        | 379.00        |                |
| <b>TOTAL</b>   | <b>37.00</b>  | <b>172.00</b> | <b>13.00</b>  | <b>139.00</b> | <b>39.00</b> | <b>39.00</b> | <b>439.00</b> |                |
| Exp.           | 5.06          | 23.51         | 1.78          | 19.00         | 5.33         | 5.33         |               | <b>p-value</b> |
| Exp.           | 31.94         | 148.49        | 11.22         | 120.00        | 33.67        | 33.67        |               | 9.88894E-19    |
| X <sup>2</sup> | 0.18          | 13.04         | 58.82         | 0.21          | 0.52         | 8.35         | <b>81.12</b>  |                |
| X <sup>2</sup> | 0.03          | 2.06          | 9.31          | 0.03          | 0.08         | 1.32         | <b>12.84</b>  |                |
| ACC            | 51.00         | 7.00          | 4.00          | 3.00          | 5.00         | 2.00         | 72.00         |                |
| ACC            | 77.00         | 41.00         | 13.00         | 28.00         | 8.00         | 34.00        | 201.00        |                |
| <b>TOTAL</b>   | <b>128.00</b> | <b>48.00</b>  | <b>17.00</b>  | <b>31.00</b>  | <b>13.00</b> | <b>36.00</b> | <b>273.00</b> |                |
| Exp.           | 33.76         | 12.66         | 4.48          | 8.18          | 3.43         | 9.49         |               | <b>p-value</b> |
| Exp.           | 94.24         | 35.34         | 12.52         | 22.82         | 9.57         | 26.51        |               | 2.39207E-05    |
| X <sup>2</sup> | 8.81          | 2.53          | 0.05          | 3.28          | 0.72         | 5.92         | <b>21.30</b>  |                |
| X <sup>2</sup> | 3.15          | 0.91          | 0.02          | 1.17          | 0.26         | 2.12         | <b>7.63</b>   |                |
| ACT            | 9.00          | 13.00         | 9.00          | 9.00          | 7.00         | 5.00         | 52.00         |                |
| ACT            | 13.00         | 14.00         | 18.00         | 49.00         | 0.00         | 4.00         | 98.00         |                |
| <b>TOTAL</b>   | <b>22.00</b>  | <b>27.00</b>  | <b>27.00</b>  | <b>58.00</b>  | <b>7.00</b>  | <b>9.00</b>  | <b>150.00</b> |                |
| Exp.           | 7.63          | 9.36          | 9.36          | 20.11         | 2.43         | 3.12         |               | <b>p-value</b> |
| Exp.           | 14.37         | 17.64         | 17.64         | 37.89         | 4.57         | 5.88         |               | 6.01095E-05    |
| X <sup>2</sup> | 0.25          | 1.42          | 0.01          | 6.14          | 8.62         | 1.13         | <b>17.56</b>  |                |
| X <sup>2</sup> | 0.13          | 0.75          | 0.01          | 3.26          | 4.57         | 0.60         | <b>9.32</b>   |                |
| ACG            | 4.00          | 8.00          | 5.00          | 10.00         | 6.00         | 8.00         | 41.00         |                |
| ACG            | 0.00          | 1.00          | 0.00          | 60.00         | 0.00         | 27.00        | 88.00         |                |
| <b>TOTAL</b>   | <b>4.00</b>   | <b>9.00</b>   | <b>5.00</b>   | <b>70.00</b>  | <b>6.00</b>  | <b>35.00</b> | <b>129.00</b> |                |
| Exp.           | 1.27          | 2.86          | 1.59          | 22.25         | 1.91         | 11.12        |               | <b>p-value</b> |
| Exp.           | 2.73          | 6.14          | 3.41          | 47.75         | 4.09         | 23.88        |               | 5.29652E-11*   |
| X <sup>2</sup> | 5.86          | 9.23          | 7.32          | 6.74          | 8.79         | 0.88         | <b>38.82</b>  |                |



|              |               |              |               |               |               |               |               |                |
|--------------|---------------|--------------|---------------|---------------|---------------|---------------|---------------|----------------|
| X^2          | 2.73          | 4.30         | 3.41          | 3.14          | 4.09          | 0.41          | <b>18.09</b>  |                |
| AAT          | 9.00          | 16.00        | 11.00         | 79.00         | 16.00         | 25.00         | 156.00        |                |
| AAT          | 8.00          | 40.00        | 2.00          | 32.00         | 81.00         | 140.00        | 303.00        |                |
| <b>TOTAL</b> | <b>17.00</b>  | <b>56.00</b> | <b>13.00</b>  | <b>111.00</b> | <b>97.00</b>  | <b>165.00</b> | <b>459.00</b> |                |
| Exp.         | 5.78          | 19.03        | 4.42          | 37.73         | 32.97         | 56.08         |               | <b>p-value</b> |
| Exp.         | 11.22         | 36.97        | 8.58          | 73.27         | 64.03         | 108.92        |               | 1.65255E-25    |
| X^2          | 1.80          | 0.48         | 9.80          | 45.16         | 8.73          | 17.22         | <b>83.20</b>  |                |
| X^2          | 0.93          | 0.25         | 5.05          | 23.25         | 4.50          | 8.87          | <b>42.83</b>  |                |
| ATA          | 22.00         | 14.00        | 11.00         | 10.00         | 13.00         | 6.00          | 76.00         |                |
| ATA          | 70.00         | 10.00        | 71.00         | 11.00         | 34.00         | 22.00         | 218.00        |                |
| <b>TOTAL</b> | <b>92.00</b>  | <b>24.00</b> | <b>82.00</b>  | <b>21.00</b>  | <b>47.00</b>  | <b>28.00</b>  | <b>294.00</b> |                |
| Exp.         | 23.78         | 6.20         | 21.20         | 5.43          | 12.15         | 7.24          |               | <b>p-value</b> |
| Exp.         | 68.22         | 17.80        | 60.80         | 15.57         | 34.85         | 20.76         |               | 0.000108357    |
| X^2          | 0.13          | 9.80         | 4.91          | 3.85          | 0.06          | 0.21          | <b>18.96</b>  |                |
| X^2          | 0.05          | 3.42         | 1.71          | 1.34          | 0.02          | 0.07          | <b>6.61</b>   |                |
| ATT          | 79.00         | 19.00        | 41.00         | 17.00         | 30.00         | 12.00         | 198.00        |                |
| ATT          | 30.00         | 10.00        | 53.00         | 4.00          | 156.00        | 96.00         | 349.00        |                |
| <b>TOTAL</b> | <b>109.00</b> | <b>29.00</b> | <b>94.00</b>  | <b>21.00</b>  | <b>186.00</b> | <b>108.00</b> | <b>547.00</b> |                |
| Exp.         | 39.46         | 10.50        | 34.03         | 7.60          | 67.33         | 39.09         |               | <b>p-value</b> |
| Exp.         | 69.54         | 18.50        | 59.97         | 13.40         | 118.67        | 68.91         |               | 1.02567E-31    |
| X^2          | 39.63         | 6.89         | 1.43          | 11.62         | 20.69         | 18.78         | <b>99.04</b>  |                |
| X^2          | 22.49         | 3.91         | 0.81          | 6.59          | 11.74         | 10.65         | <b>56.19</b>  |                |
| ATC          | 17.00         | 10.00        | 19.00         | 12.00         | 10.00         | 7.00          | 75.00         |                |
| ATC          | 1.00          | 6.00         | 0.00          | 81.00         | 0.00          | 62.00         | 150.00        |                |
| <b>TOTAL</b> | <b>18.00</b>  | <b>16.00</b> | <b>19.00</b>  | <b>93.00</b>  | <b>10.00</b>  | <b>69.00</b>  | <b>225.00</b> |                |
| Exp.         | 6.00          | 5.33         | 6.33          | 31.00         | 3.33          | 23.00         |               | <b>p-value</b> |
| Exp.         | 12.00         | 10.67        | 12.67         | 62.00         | 6.67          | 46.00         |               | 4.86079E-26    |
| X^2          | 20.17         | 4.08         | 25.33         | 11.65         | 13.33         | 11.13         | <b>85.69</b>  |                |
| X^2          | 10.08         | 2.04         | 12.67         | 5.82          | 6.67          | 5.57          | <b>42.85</b>  |                |
| ATG          | 10.00         | 9.00         | 25.00         | 12.00         | 70.00         | 18.00         | 144.00        |                |
| ATG          | 79.00         | 1.00         | 35.00         | 0.00          | 0.00          | 0.00          | 115.00        |                |
| <b>TOTAL</b> | <b>89.00</b>  | <b>10.00</b> | <b>60.00</b>  | <b>12.00</b>  | <b>70.00</b>  | <b>18.00</b>  | <b>259.00</b> |                |
| Exp.         | 49.48         | 5.56         | 33.36         | 6.67          | 38.92         | 10.01         |               | <b>p-value</b> |
| Exp.         | 39.52         | 4.44         | 26.64         | 5.33          | 31.08         | 7.99          |               | 8.44338E-33    |
| X^2          | 31.50         | 2.13         | 2.09          | 4.26          | 24.82         | 6.38          | <b>71.19</b>  |                |
| X^2          | 39.45         | 2.67         | 2.62          | 5.33          | 31.08         | 7.99          | <b>89.14</b>  |                |
| AAG          | 55.00         | 21.00        | 58.00         | 7.00          | 18.00         | 16.00         | 175.00        |                |
| AAG          | 8.00          | 4.00         | 13.00         | 17.00         | 4.00          | 8.00          | 54.00         |                |
| <b>TOTAL</b> | <b>63.00</b>  | <b>25.00</b> | <b>71.00</b>  | <b>24.00</b>  | <b>22.00</b>  | <b>24.00</b>  | <b>229.00</b> |                |
| Exp.         | 48.14         | 19.10        | 54.26         | 18.34         | 16.81         | 18.34         |               | <b>p-value</b> |
| Exp.         | 14.86         | 5.90         | 16.74         | 5.66          | 5.19          | 5.66          |               | 4.9977E-07     |
| X^2          | 0.98          | 0.19         | 0.26          | 7.01          | 0.08          | 0.30          | <b>8.82</b>   |                |
| X^2          | 3.16          | 0.61         | 0.84          | 22.72         | 0.27          | 0.97          | <b>28.57</b>  |                |
| AGA          | 7.00          | 65.00        | 16.00         | 66.00         | 8.00          | 16.00         | 178.00        |                |
| AGA          | 9.00          | 0.00         | 91.00         | 0.00          | 0.00          | 2.00          | 102.00        |                |
| <b>TOTAL</b> | <b>16.00</b>  | <b>65.00</b> | <b>107.00</b> | <b>66.00</b>  | <b>8.00</b>   | <b>18.00</b>  | <b>280.00</b> |                |
| Exp.         | 10.17         | 41.32        | 68.02         | 41.96         | 5.09          | 11.44         |               | <b>p-value</b> |
| Exp.         | 5.83          | 23.68        | 38.98         | 24.04         | 2.91          | 6.56          |               | 1.54464E-40    |
| X^2          | 0.99          | 13.57        | 39.78         | 13.78         | 1.67          | 1.81          | <b>71.60</b>  |                |
| X^2          | 1.73          | 23.68        | 69.43         | 24.04         | 2.91          | 3.17          | <b>124.96</b> |                |
| AGG          | 11.00         | 37.00        | 8.00          | 48.00         | 10.00         | 38.00         | 152.00        |                |
| AGG          | 9.00          | 0.00         | 49.00         | 0.00          | 0.00          | 0.00          | 58.00         |                |
| <b>TOTAL</b> | <b>20.00</b>  | <b>37.00</b> | <b>57.00</b>  | <b>48.00</b>  | <b>10.00</b>  | <b>38.00</b>  | <b>210.00</b> |                |
| Exp.         | 14.48         | 26.78        | 41.26         | 34.74         | 7.24          | 27.50         |               | <b>p-value</b> |
| Exp.         | 5.52          | 10.22        | 15.74         | 13.26         | 2.76          | 10.50         |               | 8.85743E-31    |
| X^2          | 0.83          | 3.90         | 26.81         | 5.06          | 1.05          | 4.00          | <b>41.66</b>  |                |
| X^2          | 2.19          | 10.22        | 70.26         | 13.26         | 2.76          | 10.50         | <b>109.18</b> |                |
| AGC          | 8.00          | 18.00        | 11.00         | 88.00         | 9.00          | 16.00         | 150.00        |                |
| AGC          | 0.00          | 0.00         | 84.00         | 2.00          | 37.00         | 42.00         | 165.00        |                |
| <b>TOTAL</b> | <b>8.00</b>   | <b>18.00</b> | <b>95.00</b>  | <b>90.00</b>  | <b>46.00</b>  | <b>58.00</b>  | <b>315.00</b> |                |
| Exp.         | 3.81          | 8.57         | 45.24         | 42.86         | 21.90         | 27.62         |               | <b>p-value</b> |
| Exp.         | 4.19          | 9.43         | 49.76         | 47.14         | 24.10         | 30.38         |               | 1.03736E-39    |
| X^2          | 4.61          | 10.37        | 25.91         | 47.55         | 7.60          | 4.89          | <b>100.93</b> |                |
| X^2          | 4.19          | 9.43         | 23.56         | 43.23         | 6.91          | 4.44          | <b>91.76</b>  |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|              |              |              |              |               |               |              |               |                |
|--------------|--------------|--------------|--------------|---------------|---------------|--------------|---------------|----------------|
| AGT          | 13.00        | 8.00         | 11.00        | 4.00          | 10.00         | 8.00         | 54.00         |                |
| AGT          | 2.00         | 10.00        | 6.00         | 12.00         | 37.00         | 1.00         | 68.00         |                |
| <b>TOTAL</b> | <b>15.00</b> | <b>18.00</b> | <b>17.00</b> | <b>16.00</b>  | <b>47.00</b>  | <b>9.00</b>  | <b>122.00</b> |                |
| Exp.         | 6.64         | 7.97         | 7.52         | 7.08          | 20.80         | 3.98         |               | <b>p-value</b> |
| Exp.         | 8.36         | 10.03        | 9.48         | 8.92          | 26.20         | 5.02         |               | 2.9264E-06     |
| X^2          | 6.09         | 0.00         | 1.61         | 1.34          | 5.61          | 4.05         | <b>18.70</b>  |                |
| X^2          | 4.84         | 0.00         | 1.27         | 1.07          | 4.46          | 3.22         | <b>14.85</b>  |                |
| CAA          | 45.00        | 12.00        | 8.00         | 4.00          | 14.00         | 17.00        | 100.00        |                |
| CAA          | 0.00         | 51.00        | 0.00         | 13.00         | 68.00         | 64.00        | 196.00        |                |
| <b>TOTAL</b> | <b>45.00</b> | <b>63.00</b> | <b>8.00</b>  | <b>17.00</b>  | <b>82.00</b>  | <b>81.00</b> | <b>296.00</b> |                |
| Exp.         | 15.20        | 21.28        | 2.70         | 5.74          | 27.70         | 27.36        |               | <b>p-value</b> |
| Exp.         | 29.80        | 41.72        | 5.30         | 11.26         | 54.30         | 53.64        |               | 1.05113E-25    |
| X^2          | 58.40        | 4.05         | 10.38        | 0.53          | 6.78          | 3.93         | <b>84.07</b>  |                |
| X^2          | 29.80        | 2.07         | 5.30         | 0.27          | 3.46          | 2.00         | <b>42.89</b>  |                |
| CAC          | 5.00         | 14.00        | 14.00        | 8.00          | 48.00         | 9.00         | 98.00         |                |
| CAC          | 49.00        | 45.00        | 85.00        | 0.00          | 98.00         | 24.00        | 301.00        |                |
| <b>TOTAL</b> | <b>54.00</b> | <b>59.00</b> | <b>99.00</b> | <b>8.00</b>   | <b>146.00</b> | <b>33.00</b> | <b>399.00</b> |                |
| Exp.         | 13.26        | 14.49        | 24.32        | 1.96          | 35.86         | 8.11         |               | <b>p-value</b> |
| Exp.         | 40.74        | 44.51        | 74.68        | 6.04          | 110.14        | 24.89        |               | 4.0599E-08     |
| X^2          | 5.15         | 0.02         | 4.38         | 18.54         | 4.11          | 0.10         | <b>32.29</b>  |                |
| X^2          | 1.68         | 0.01         | 1.42         | 6.04          | 1.34          | 0.03         | <b>10.51</b>  |                |
| CCA          | 9.00         | 9.00         | 6.00         | 15.00         | 6.00          | 53.00        | 98.00         |                |
| CCA          | 7.00         | 0.00         | 34.00        | 108.00        | 16.00         | 42.00        | 207.00        |                |
| <b>TOTAL</b> | <b>16.00</b> | <b>9.00</b>  | <b>40.00</b> | <b>123.00</b> | <b>22.00</b>  | <b>95.00</b> | <b>305.00</b> |                |
| Exp.         | 5.14         | 2.89         | 12.85        | 39.52         | 7.07          | 30.52        |               | <b>p-value</b> |
| Exp.         | 10.86        | 6.11         | 27.15        | 83.48         | 14.93         | 64.48        |               | 6.64381E-15    |
| X^2          | 2.90         | 12.90        | 3.65         | 15.21         | 0.16          | 16.55        | <b>51.38</b>  |                |
| X^2          | 1.37         | 6.11         | 1.73         | 7.20          | 0.08          | 7.83         | <b>24.32</b>  |                |
| CCC          | 5.00         | 17.00        | 3.00         | 15.00         | 4.00          | 15.00        | 59.00         |                |
| CCC          | 0.00         | 5.00         | 0.00         | 91.00         | 0.00          | 42.00        | 138.00        |                |
| <b>TOTAL</b> | <b>5.00</b>  | <b>22.00</b> | <b>3.00</b>  | <b>106.00</b> | <b>4.00</b>   | <b>57.00</b> | <b>197.00</b> |                |
| Exp.         | 1.50         | 6.59         | 0.90         | 31.75         | 1.20          | 17.07        |               | <b>p-value</b> |
| Exp.         | 3.50         | 15.41        | 2.10         | 74.25         | 2.80          | 39.93        |               | 1.40869E-12*   |
| X^2          | 8.19         | 16.45        | 4.92         | 8.83          | 6.55          | 0.25         | <b>45.20</b>  |                |
| X^2          | 3.50         | 7.03         | 2.10         | 3.78          | 2.80          | 0.11         | <b>19.32</b>  |                |
| CCT          | 9.00         | 22.00        | 13.00        | 39.00         | 7.00          | 13.00        | 103.00        |                |
| CCT          | 15.00        | 36.00        | 6.00         | 15.00         | 39.00         | 76.00        | 187.00        |                |
| <b>TOTAL</b> | <b>24.00</b> | <b>58.00</b> | <b>19.00</b> | <b>54.00</b>  | <b>46.00</b>  | <b>89.00</b> | <b>290.00</b> |                |
| Exp.         | 8.52         | 20.60        | 6.75         | 19.18         | 16.34         | 31.61        |               | <b>p-value</b> |
| Exp.         | 15.48        | 37.40        | 12.25        | 34.82         | 29.66         | 57.39        |               | 6.30199E-13    |
| X^2          | 0.03         | 0.10         | 5.79         | 20.48         | 5.34          | 10.96        | <b>42.69</b>  |                |
| X^2          | 0.01         | 0.05         | 3.19         | 11.28         | 2.94          | 6.03         | <b>23.51</b>  |                |
| CCG          | 1.00         | 4.00         | 6.00         | 12.00         | 6.00          | 7.00         | 36.00         |                |
| CCG          | 3.00         | 32.00        | 16.00        | 0.00          | 2.00          | 32.00        | 85.00         |                |
| <b>TOTAL</b> | <b>4.00</b>  | <b>36.00</b> | <b>22.00</b> | <b>12.00</b>  | <b>8.00</b>   | <b>39.00</b> | <b>121.00</b> |                |
| Exp.         | 1.19         | 10.71        | 6.55         | 3.57          | 2.38          | 11.60        |               | <b>p-value</b> |
| Exp.         | 2.81         | 25.29        | 15.45        | 8.43          | 5.62          | 27.40        |               | 1.54692E-08*   |
| X^2          | 0.03         | 4.20         | 0.05         | 19.90         | 5.51          | 1.83         | <b>31.52</b>  |                |
| X^2          | 0.01         | 1.78         | 0.02         | 8.43          | 2.33          | 0.77         | <b>13.35</b>  |                |
| CAT          | 11.00        | 10.00        | 11.00        | 8.00          | 14.00         | 8.00         | 62.00         |                |
| CAT          | 36.00        | 84.00        | 21.00        | 3.00          | 66.00         | 21.00        | 231.00        |                |
| <b>TOTAL</b> | <b>47.00</b> | <b>94.00</b> | <b>32.00</b> | <b>11.00</b>  | <b>80.00</b>  | <b>29.00</b> | <b>293.00</b> |                |
| Exp.         | 9.95         | 19.89        | 6.77         | 2.33          | 16.93         | 6.14         |               | <b>p-value</b> |
| Exp.         | 37.05        | 74.11        | 25.23        | 8.67          | 63.07         | 22.86        |               | 2.74849E-05    |
| X^2          | 0.11         | 4.92         | 2.64         | 13.82         | 0.51          | 0.57         | <b>22.57</b>  |                |
| X^2          | 0.03         | 1.32         | 0.71         | 3.71          | 0.14          | 0.15         | <b>6.06</b>   |                |
| CTA          | 3.00         | 7.00         | 5.00         | 11.00         | 8.00          | 7.00         | 41.00         |                |
| CTA          | 20.00        | 2.00         | 40.00        | 16.00         | 45.00         | 9.00         | 132.00        |                |
| <b>TOTAL</b> | <b>23.00</b> | <b>9.00</b>  | <b>45.00</b> | <b>27.00</b>  | <b>53.00</b>  | <b>16.00</b> | <b>173.00</b> |                |
| Exp.         | 5.45         | 2.13         | 10.66        | 6.40          | 12.56         | 3.79         |               | <b>p-value</b> |
| Exp.         | 17.55        | 6.87         | 34.34        | 20.60         | 40.44         | 12.21        |               | 1.47031E-05    |
| X^2          | 1.10         | 11.11        | 3.01         | 3.31          | 1.66          | 2.71         | <b>22.90</b>  |                |
| X^2          | 0.34         | 3.45         | 0.93         | 1.03          | 0.51          | 0.84         | <b>7.11</b>   |                |

|              |               |               |              |               |              |               |               |                |
|--------------|---------------|---------------|--------------|---------------|--------------|---------------|---------------|----------------|
| CTT          | 24.00         | 23.00         | 14.00        | 11.00         | 9.00         | 9.00          | 90.00         |                |
| CTT          | 56.00         | 47.00         | 24.00        | 0.00          | 35.00        | 32.00         | 194.00        |                |
| <b>TOTAL</b> | <b>80.00</b>  | <b>70.00</b>  | <b>38.00</b> | <b>11.00</b>  | <b>44.00</b> | <b>41.00</b>  | <b>284.00</b> |                |
| Exp.         | 25.35         | 22.18         | 12.04        | 3.49          | 13.94        | 12.99         |               | <b>p-value</b> |
| Exp.         | 54.65         | 47.82         | 25.96        | 7.51          | 30.06        | 28.01         |               | 2.6684E-05     |
| X^2          | 0.07          | 0.03          | 0.32         | 16.20         | 1.75         | 1.23          | <b>19.60</b>  |                |
| X^2          | 0.03          | 0.01          | 0.15         | 7.51          | 0.81         | 0.57          | <b>9.09</b>   |                |
| CTC          | 6.00          | 10.00         | 11.00        | 14.00         | 15.00        | 8.00          | 64.00         |                |
| CTC          | 0.00          | 12.00         | 1.00         | 24.00         | 0.00         | 24.00         | 61.00         |                |
| <b>TOTAL</b> | <b>6.00</b>   | <b>22.00</b>  | <b>12.00</b> | <b>38.00</b>  | <b>15.00</b> | <b>32.00</b>  | <b>125.00</b> |                |
| Exp.         | 3.07          | 11.26         | 6.14         | 19.46         | 7.68         | 16.38         |               | <b>p-value</b> |
| Exp.         | 2.93          | 10.74         | 5.86         | 18.54         | 7.32         | 15.62         |               | 1.42705E-07    |
| X^2          | 2.79          | 0.14          | 3.84         | 1.53          | 6.98         | 4.29          | <b>19.57</b>  |                |
| X^2          | 2.93          | 0.15          | 4.03         | 1.61          | 7.32         | 4.50          | <b>20.53</b>  |                |
| CTG          | 4.00          | 5.00          | 4.00         | 10.00         | 5.00         | 53.00         | 81.00         |                |
| CTG          | 115.00        | 0.00          | 45.00        | 0.00          | 28.00        | 4.00          | 192.00        |                |
| <b>TOTAL</b> | <b>119.00</b> | <b>5.00</b>   | <b>49.00</b> | <b>10.00</b>  | <b>33.00</b> | <b>57.00</b>  | <b>273.00</b> |                |
| Exp.         | 35.31         | 1.48          | 14.54        | 2.97          | 9.79         | 16.91         |               | <b>p-value</b> |
| Exp.         | 83.69         | 3.52          | 34.46        | 7.03          | 23.21        | 40.09         |               | 5.34485E-41    |
| X^2          | 27.76         | 8.34          | 7.64         | 16.67         | 2.34         | 77.01         | <b>139.76</b> |                |
| X^2          | 11.71         | 3.52          | 3.22         | 7.03          | 0.99         | 32.49         | <b>58.96</b>  |                |
| CAG          | 17.00         | 14.00         | 54.00        | 5.00          | 11.00        | 8.00          | 109.00        |                |
| CAG          | 2.00          | 2.00          | 0.00         | 35.00         | 8.00         | 139.00        | 186.00        |                |
| <b>TOTAL</b> | <b>19.00</b>  | <b>16.00</b>  | <b>54.00</b> | <b>40.00</b>  | <b>19.00</b> | <b>147.00</b> | <b>295.00</b> |                |
| Exp.         | 7.02          | 5.91          | 19.95        | 14.78         | 7.02         | 54.32         |               | <b>p-value</b> |
| Exp.         | 11.98         | 10.09         | 34.05        | 25.22         | 11.98        | 92.68         |               | 3.9521E-43     |
| X^2          | 14.19         | 11.07         | 58.10        | 6.47          | 2.26         | 39.49         | <b>131.57</b> |                |
| X^2          | 8.31          | 6.48          | 34.05        | 3.79          | 1.32         | 23.14         | <b>77.10</b>  |                |
| CGA          | 3.00          | 8.00          | 6.00         | 6.00          | 10.00        | 10.00         | 43.00         |                |
| CGA          | 3.00          | 0.00          | 61.00        | 0.00          | 25.00        | 0.00          | 89.00         |                |
| <b>TOTAL</b> | <b>6.00</b>   | <b>8.00</b>   | <b>67.00</b> | <b>6.00</b>   | <b>35.00</b> | <b>10.00</b>  | <b>132.00</b> |                |
| Exp.         | 1.95          | 2.61          | 21.83        | 1.95          | 11.40        | 3.26          |               | <b>p-value</b> |
| Exp.         | 4.05          | 5.39          | 45.17        | 4.05          | 23.60        | 6.74          |               | 2.96881E-13*   |
| X^2          | 0.56          | 11.16         | 11.48        | 8.37          | 0.17         | 13.96         | <b>45.70</b>  |                |
| X^2          | 0.27          | 5.39          | 5.54         | 4.05          | 0.08         | 6.74          | <b>22.08</b>  |                |
| CGG          | 19.00         | 39.00         | 3.00         | 10.00         | 5.00         | 8.00          | 84.00         |                |
| CGG          | 11.00         | 2.00          | 1.00         | 0.00          | 31.00        | 0.00          | 45.00         |                |
| <b>TOTAL</b> | <b>30.00</b>  | <b>41.00</b>  | <b>4.00</b>  | <b>10.00</b>  | <b>36.00</b> | <b>8.00</b>   | <b>129.00</b> |                |
| Exp.         | 19.53         | 26.70         | 2.60         | 6.51          | 23.44        | 5.21          |               | <b>p-value</b> |
| Exp.         | 10.47         | 14.30         | 1.40         | 3.49          | 12.56        | 2.79          |               | 3.08516E-13*   |
| X^2          | 0.01          | 5.67          | 0.06         | 1.87          | 14.51        | 1.50          | <b>23.62</b>  |                |
| X^2          | 0.03          | 10.58         | 0.11         | 3.49          | 27.08        | 2.79          | <b>44.08</b>  |                |
| CGC          | 4.00          | 6.00          | 10.00        | 3.00          | 11.00        | 11.00         | 45.00         |                |
| CGC          | 1.00          | 1.00          | 53.00        | 52.00         | 32.00        | 33.00         | 172.00        |                |
| <b>TOTAL</b> | <b>5.00</b>   | <b>7.00</b>   | <b>63.00</b> | <b>55.00</b>  | <b>43.00</b> | <b>44.00</b>  | <b>217.00</b> |                |
| Exp.         | 1.04          | 1.45          | 13.06        | 11.41         | 8.92         | 9.12          |               | <b>p-value</b> |
| Exp.         | 3.96          | 5.55          | 49.94        | 43.59         | 34.08        | 34.88         |               | 3.01308E-07    |
| X^2          | 8.47          | 14.25         | 0.72         | 6.19          | 0.49         | 0.39          | <b>30.51</b>  |                |
| X^2          | 2.22          | 3.73          | 0.19         | 1.62          | 0.13         | 0.10          | <b>7.98</b>   |                |
| CGT          | 51.00         | 1.00          | 11.00        | 5.00          | 2.00         | 9.00          | 79.00         |                |
| CGT          | 0.00          | 0.00          | 21.00        | 0.00          | 21.00        | 0.00          | 42.00         |                |
| <b>TOTAL</b> | <b>51.00</b>  | <b>1.00</b>   | <b>32.00</b> | <b>5.00</b>   | <b>23.00</b> | <b>9.00</b>   | <b>121.00</b> |                |
| Exp.         | 33.30         | 0.65          | 20.89        | 3.26          | 15.02        | 5.88          |               | <b>p-value</b> |
| Exp.         | 17.70         | 0.35          | 11.11        | 1.74          | 7.98         | 3.12          |               | 4.96619E-16*   |
| X^2          | 9.41          | 0.18          | 4.68         | 0.92          | 11.28        | 1.66          | <b>28.15</b>  |                |
| X^2          | 17.70         | 0.35          | 8.81         | 1.74          | 21.22        | 3.12          | <b>52.94</b>  |                |
| TAA          | 8.00          | 12.00         | 6.00         | 14.00         | 15.00        | 8.00          | 63.00         |                |
| TAA          | 35.00         | 110.00        | 5.00         | 121.00        | 0.00         | 37.00         | 308.00        |                |
| <b>TOTAL</b> | <b>43.00</b>  | <b>122.00</b> | <b>11.00</b> | <b>135.00</b> | <b>15.00</b> | <b>45.00</b>  | <b>371.00</b> |                |
| Exp.         | 7.30          | 20.72         | 1.87         | 22.92         | 2.55         | 7.64          |               | <b>p-value</b> |
| Exp.         | 35.70         | 101.28        | 9.13         | 112.08        | 12.45        | 37.36         |               | 1.53755E-18    |
| X^2          | 0.07          | 3.67          | 9.14         | 3.47          | 60.88        | 0.02          | <b>77.25</b>  |                |
| X^2          | 0.01          | 0.75          | 1.87         | 0.71          | 12.45        | 0.00          | <b>15.80</b>  |                |
| TAC          | 5.00          | 5.00          | 4.00         | 4.00          | 9.00         | 8.00          | 35.00         |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|       |        |        |        |        |        |        |               |                |
|-------|--------|--------|--------|--------|--------|--------|---------------|----------------|
| TAC   | 70.00  | 2.00   | 58.00  | 1.00   | 3.00   | 1.00   | 135.00        |                |
| TOTAL | 75.00  | 7.00   | 62.00  | 5.00   | 12.00  | 9.00   | 170.00        |                |
| Exp.  | 15.44  | 1.44   | 12.76  | 1.03   | 2.47   | 1.85   |               | <b>p-value</b> |
| Exp.  | 59.56  | 5.56   | 49.24  | 3.97   | 9.53   | 7.15   |               | 5.26601E-17*   |
| X^2   | 7.06   | 8.79   | 6.02   | 8.57   | 17.26  | 20.39  | <b>68.09</b>  |                |
| X^2   | 1.83   | 2.28   | 1.56   | 2.22   | 4.47   | 5.29   | <b>17.65</b>  |                |
| TCA   | 8.00   | 51.00  | 12.00  | 24.00  | 14.00  | 9.00   | 118.00        |                |
| TCA   | 14.00  | 0.00   | 67.00  | 0.00   | 34.00  | 0.00   | 115.00        |                |
| TOTAL | 22.00  | 51.00  | 79.00  | 24.00  | 48.00  | 9.00   | 233.00        |                |
| Exp.  | 11.14  | 25.83  | 40.01  | 12.15  | 24.31  | 4.56   |               | <b>p-value</b> |
| Exp.  | 10.86  | 25.17  | 38.99  | 11.85  | 23.69  | 4.44   |               | 7.94784E-27    |
| X^2   | 0.89   | 24.53  | 19.61  | 11.54  | 4.37   | 4.33   | <b>65.27</b>  |                |
| X^2   | 0.91   | 25.17  | 20.12  | 11.85  | 4.49   | 4.44   | <b>66.97</b>  |                |
| TCC   | 18.00  | 53.00  | 16.00  | 6.00   | 13.00  | 12.00  | 118.00        |                |
| TCC   | 3.00   | 27.00  | 18.00  | 21.00  | 56.00  | 14.00  | 139.00        |                |
| TOTAL | 21.00  | 80.00  | 34.00  | 27.00  | 69.00  | 26.00  | 257.00        |                |
| Exp.  | 9.64   | 36.73  | 15.61  | 12.40  | 31.68  | 11.94  |               | <b>p-value</b> |
| Exp.  | 11.36  | 43.27  | 18.39  | 14.60  | 37.32  | 14.06  |               | 3.05193E-10    |
| X^2   | 7.24   | 7.21   | 0.01   | 3.30   | 11.02  | 0.00   | <b>28.78</b>  |                |
| X^2   | 6.15   | 6.12   | 0.01   | 2.80   | 9.35   | 0.00   | <b>24.43</b>  |                |
| TCT   | 15.00  | 9.00   | 21.00  | 17.00  | 26.00  | 8.00   | 96.00         |                |
| TCT   | 17.00  | 23.00  | 57.00  | 4.00   | 79.00  | 0.00   | 180.00        |                |
| TOTAL | 32.00  | 32.00  | 78.00  | 21.00  | 105.00 | 8.00   | 276.00        |                |
| Exp.  | 11.13  | 11.13  | 27.13  | 7.30   | 36.52  | 2.78   |               | <b>p-value</b> |
| Exp.  | 20.87  | 20.87  | 50.87  | 13.70  | 68.48  | 5.22   |               | 2.11567E-08    |
| X^2   | 1.35   | 0.41   | 1.39   | 12.87  | 3.03   | 9.78   | <b>28.82</b>  |                |
| X^2   | 0.72   | 0.22   | 0.74   | 6.86   | 1.62   | 5.22   | <b>15.37</b>  |                |
| TCG   | 5.00   | 9.00   | 6.00   | 4.00   | 3.00   | 8.00   | 35.00         |                |
| TCG   | 0.00   | 0.00   | 2.00   | 1.00   | 33.00  | 42.00  | 78.00         |                |
| TOTAL | 5.00   | 9.00   | 8.00   | 5.00   | 36.00  | 50.00  | 113.00        |                |
| Exp.  | 1.55   | 2.79   | 2.48   | 1.55   | 11.15  | 15.49  |               | <b>p-value</b> |
| Exp.  | 3.45   | 6.21   | 5.52   | 3.45   | 24.85  | 34.51  |               | 3.22366E-11*   |
| X^2   | 7.69   | 13.84  | 5.01   | 3.88   | 5.96   | 3.62   | <b>40.00</b>  |                |
| X^2   | 3.45   | 6.21   | 2.25   | 1.74   | 2.67   | 1.62   | <b>17.95</b>  |                |
| TAT   | 22.00  | 53.00  | 34.00  | 22.00  | 25.00  | 21.00  | 177.00        |                |
| TAT   | 73.00  | 102.00 | 10.00  | 16.00  | 2.00   | 8.00   | 211.00        |                |
| TOTAL | 95.00  | 155.00 | 44.00  | 38.00  | 27.00  | 29.00  | 388.00        |                |
| Exp.  | 43.34  | 70.71  | 20.07  | 17.34  | 12.32  | 13.23  |               | <b>p-value</b> |
| Exp.  | 51.66  | 84.29  | 23.93  | 20.66  | 14.68  | 15.77  |               | 8.54536E-16    |
| X^2   | 10.51  | 4.44   | 9.66   | 1.26   | 13.06  | 4.56   | <b>43.48</b>  |                |
| X^2   | 8.81   | 3.72   | 8.11   | 1.05   | 10.96  | 3.83   | <b>36.48</b>  |                |
| TTA   | 10.00  | 37.00  | 6.00   | 42.00  | 21.00  | 27.00  | 143.00        |                |
| TTA   | 104.00 | 26.00  | 113.00 | 5.00   | 9.00   | 13.00  | 270.00        |                |
| TOTAL | 114.00 | 63.00  | 119.00 | 47.00  | 30.00  | 40.00  | 413.00        |                |
| Exp.  | 39.47  | 21.81  | 41.20  | 16.27  | 10.39  | 13.85  |               | <b>p-value</b> |
| Exp.  | 74.53  | 41.19  | 77.80  | 30.73  | 19.61  | 26.15  |               | 6.21845E-40    |
| X^2   | 22.01  | 10.57  | 30.08  | 40.67  | 10.84  | 12.49  | <b>126.65</b> |                |
| X^2   | 11.65  | 5.60   | 15.93  | 21.54  | 5.74   | 6.61   | <b>67.08</b>  |                |
| TTT   | 96.00  | 101.00 | 46.00  | 24.00  | 17.00  | 19.00  | 303.00        |                |
| TTT   | 1.00   | 41.00  | 0.00   | 81.00  | 0.00   | 71.00  | 194.00        |                |
| TOTAL | 97.00  | 142.00 | 46.00  | 105.00 | 17.00  | 90.00  | 497.00        |                |
| Exp.  | 59.14  | 86.57  | 28.04  | 64.01  | 10.36  | 54.87  |               | <b>p-value</b> |
| Exp.  | 37.86  | 55.43  | 17.96  | 40.99  | 6.64   | 35.13  |               | 1.35895E-47    |
| X^2   | 22.98  | 2.40   | 11.50  | 25.01  | 4.25   | 23.45  | <b>89.59</b>  |                |
| X^2   | 35.89  | 3.76   | 17.96  | 39.07  | 6.64   | 36.62  | <b>139.93</b> |                |
| TTC   | 15.00  | 13.00  | 45.00  | 32.00  | 20.00  | 17.00  | 142.00        |                |
| TTC   | 49.00  | 0.00   | 1.00   | 12.00  | 14.00  | 115.00 | 191.00        |                |
| TOTAL | 64.00  | 13.00  | 46.00  | 44.00  | 34.00  | 132.00 | 333.00        |                |
| Exp.  | 27.29  | 5.54   | 19.62  | 18.76  | 14.50  | 56.29  |               | <b>p-value</b> |
| Exp.  | 36.71  | 7.46   | 26.38  | 25.24  | 19.50  | 75.71  |               | 4.67497E-31    |
| X^2   | 5.54   | 10.03  | 32.85  | 9.34   | 2.09   | 27.42  | <b>87.26</b>  |                |
| X^2   | 4.12   | 7.46   | 24.42  | 6.94   | 1.55   | 20.39  | <b>64.88</b>  |                |
| TTG   | 6.00   | 16.00  | 9.00   | 26.00  | 20.00  | 77.00  | 154.00        |                |
| TTG   | 62.00  | 0.00   | 87.00  | 0.00   | 0.00   | 0.00   | 149.00        |                |

|              |       |       |       |        |       |       |               |                |
|--------------|-------|-------|-------|--------|-------|-------|---------------|----------------|
| <b>TOTAL</b> | 68.00 | 16.00 | 96.00 | 26.00  | 20.00 | 77.00 | 303.00        |                |
| Exp.         | 34.56 | 8.13  | 48.79 | 13.21  | 10.17 | 39.14 |               | <b>p-value</b> |
| Exp.         | 33.44 | 7.87  | 47.21 | 12.79  | 9.83  | 37.86 |               | 1.16604E-51    |
| <b>X^2</b>   | 23.60 | 7.61  | 32.45 | 12.37  | 9.52  | 36.64 | <b>122.19</b> |                |
| <b>X^2</b>   | 24.39 | 7.87  | 33.54 | 12.79  | 9.83  | 37.86 | <b>126.29</b> |                |
| <b>TAG</b>   | 4.00  | 7.00  | 11.00 | 9.00   | 19.00 | 39.00 | 89.00         |                |
| <b>TAG</b>   | 0.00  | 0.00  | 1.00  | 49.00  | 2.00  | 3.00  | 55.00         |                |
| <b>TOTAL</b> | 4.00  | 7.00  | 12.00 | 58.00  | 21.00 | 42.00 | 144.00        |                |
| Exp.         | 2.47  | 4.33  | 7.42  | 35.85  | 12.98 | 25.96 |               | <b>p-value</b> |
| Exp.         | 1.53  | 2.67  | 4.58  | 22.15  | 8.02  | 16.04 |               | 1.42781E-17*   |
| <b>X^2</b>   | 0.94  | 1.65  | 1.73  | 20.11  | 2.79  | 6.55  | <b>33.78</b>  |                |
| <b>X^2</b>   | 1.53  | 2.67  | 2.80  | 32.54  | 4.52  | 10.60 | <b>54.66</b>  |                |
| <b>TGA</b>   | 5.00  | 8.00  | 6.00  | 6.00   | 17.00 | 5.00  | 47.00         |                |
| <b>TGA</b>   | 0.00  | 74.00 | 0.00  | 7.00   | 0.00  | 0.00  | 81.00         |                |
| <b>TOTAL</b> | 5.00  | 82.00 | 6.00  | 13.00  | 17.00 | 5.00  | 128.00        |                |
| Exp.         | 1.84  | 30.11 | 2.20  | 4.77   | 6.24  | 1.84  |               | <b>p-value</b> |
| Exp.         | 3.16  | 51.89 | 3.80  | 8.23   | 10.76 | 3.16  |               | 1.95185E-16*   |
| <b>X^2</b>   | 5.45  | 16.23 | 6.54  | 0.32   | 18.54 | 5.45  | <b>52.54</b>  |                |
| <b>X^2</b>   | 3.16  | 9.42  | 3.80  | 0.18   | 10.76 | 3.16  | <b>30.49</b>  |                |
| <b>TGG</b>   | 15.00 | 46.00 | 7.00  | 46.00  | 6.00  | 43.00 | 163.00        |                |
| <b>TGG</b>   | 0.00  | 33.00 | 0.00  | 44.00  | 0.00  | 4.00  | 81.00         |                |
| <b>TOTAL</b> | 15.00 | 79.00 | 7.00  | 90.00  | 6.00  | 47.00 | 244.00        |                |
| Exp.         | 10.02 | 52.77 | 4.68  | 60.12  | 4.01  | 31.40 |               | <b>p-value</b> |
| Exp.         | 4.98  | 26.23 | 2.32  | 29.88  | 1.99  | 15.60 |               | 1.93418E-07*   |
| <b>X^2</b>   | 2.47  | 0.87  | 1.15  | 3.32   | 0.99  | 4.29  | <b>13.09</b>  |                |
| <b>X^2</b>   | 4.98  | 1.75  | 2.32  | 6.68   | 1.99  | 8.63  | <b>26.35</b>  |                |
| <b>TGC</b>   | 5.00  | 7.00  | 5.00  | 8.00   | 4.00  | 6.00  | 35.00         |                |
| <b>TGC</b>   | 22.00 | 65.00 | 1.00  | 91.00  | 0.00  | 0.00  | 179.00        |                |
| <b>TOTAL</b> | 27.00 | 72.00 | 6.00  | 99.00  | 4.00  | 6.00  | 214.00        |                |
| Exp.         | 4.42  | 11.78 | 0.98  | 16.19  | 0.65  | 0.98  |               | <b>p-value</b> |
| Exp.         | 22.58 | 60.22 | 5.02  | 82.81  | 3.35  | 5.02  |               | 2.01504E-15*   |
| <b>X^2</b>   | 0.08  | 1.94  | 16.46 | 4.14   | 17.11 | 25.67 | <b>65.39</b>  |                |
| <b>X^2</b>   | 0.02  | 0.38  | 3.22  | 0.81   | 3.35  | 5.02  | <b>12.79</b>  |                |
| <b>TGT</b>   | 3.00  | 4.00  | 23.00 | 16.00  | 72.00 | 49.00 | 167.00        |                |
| <b>TGT</b>   | 0.00  | 21.00 | 0.00  | 200.00 | 0.00  | 0.00  | 221.00        |                |
| <b>TOTAL</b> | 3.00  | 25.00 | 23.00 | 216.00 | 72.00 | 49.00 | 388.00        |                |
| Exp.         | 1.29  | 10.76 | 9.90  | 92.97  | 30.99 | 21.09 |               | <b>p-value</b> |
| Exp.         | 1.71  | 14.24 | 13.10 | 123.03 | 41.01 | 27.91 |               | 1.04523E-65    |
| <b>X^2</b>   | 2.26  | 4.25  | 17.34 | 63.72  | 54.27 | 36.93 | <b>178.77</b> |                |
| <b>X^2</b>   | 1.71  | 3.21  | 13.10 | 48.15  | 41.01 | 27.91 | <b>135.09</b> |                |
| <b>GAA</b>   | 29.00 | 13.00 | 61.00 | 9.00   | 48.00 | 57.00 | 217.00        |                |
| <b>GAA</b>   | 1.00  | 43.00 | 0.00  | 2.00   | 5.00  | 34.00 | 85.00         |                |
| <b>TOTAL</b> | 30.00 | 56.00 | 61.00 | 11.00  | 53.00 | 91.00 | 302.00        |                |
| Exp.         | 21.56 | 40.24 | 43.83 | 7.90   | 38.08 | 65.39 |               | <b>p-value</b> |
| Exp.         | 8.44  | 15.76 | 17.17 | 3.10   | 14.92 | 25.61 |               | 1.49219E-22    |
| <b>X^2</b>   | 2.57  | 18.44 | 6.73  | 0.15   | 2.58  | 1.08  | <b>31.54</b>  |                |
| <b>X^2</b>   | 6.56  | 47.07 | 17.17 | 0.39   | 6.59  | 2.75  | <b>80.53</b>  |                |
| <b>GAC</b>   | 9.00  | 4.00  | 5.00  | 3.00   | 9.00  | 7.00  | 37.00         |                |
| <b>GAC</b>   | 11.00 | 57.00 | 0.00  | 28.00  | 0.00  | 1.00  | 97.00         |                |
| <b>TOTAL</b> | 20.00 | 61.00 | 5.00  | 31.00  | 9.00  | 8.00  | 134.00        |                |
| Exp.         | 5.52  | 16.84 | 1.38  | 8.56   | 2.49  | 2.21  |               | <b>p-value</b> |
| Exp.         | 14.48 | 44.16 | 3.62  | 22.44  | 6.51  | 5.79  |               | 2.94479E-14*   |
| <b>X^2</b>   | 2.19  | 9.79  | 9.49  | 3.61   | 17.08 | 10.39 | <b>52.55</b>  |                |
| <b>X^2</b>   | 0.84  | 3.74  | 3.62  | 1.38   | 6.51  | 3.96  | <b>20.05</b>  |                |
| <b>GCA</b>   | 9.00  | 6.00  | 11.00 | 3.00   | 53.00 | 10.00 | 92.00         |                |
| <b>GCA</b>   | 38.00 | 65.00 | 46.00 | 23.00  | 1.00  | 45.00 | 218.00        |                |
| <b>TOTAL</b> | 47.00 | 71.00 | 57.00 | 26.00  | 54.00 | 55.00 | 310.00        |                |
| Exp.         | 13.95 | 21.07 | 16.92 | 7.72   | 16.03 | 16.32 |               | <b>p-value</b> |
| Exp.         | 33.05 | 49.93 | 40.08 | 18.28  | 37.97 | 38.68 |               | 1.58114E-30    |
| <b>X^2</b>   | 1.76  | 10.78 | 2.07  | 2.88   | 85.31 | 2.45  | <b>105.24</b> |                |
| <b>X^2</b>   | 0.74  | 4.55  | 0.87  | 1.22   | 36.00 | 1.03  | <b>44.41</b>  |                |
| <b>GCC</b>   | 13.00 | 8.00  | 11.00 | 11.00  | 2.00  | 7.00  | 52.00         |                |
| <b>GCC</b>   | 18.00 | 32.00 | 39.00 | 0.00   | 64.00 | 1.00  | 154.00        |                |
| <b>TOTAL</b> | 31.00 | 40.00 | 50.00 | 11.00  | 66.00 | 8.00  | 206.00        |                |

Characterizing patterns in DNA sequence trace data through informatics tools

|       |       |       |       |       |       |        |              |                |
|-------|-------|-------|-------|-------|-------|--------|--------------|----------------|
| Exp.  | 7.83  | 10.10 | 12.62 | 2.78  | 16.66 | 2.02   |              | <b>p-value</b> |
| Exp.  | 23.17 | 29.90 | 37.38 | 8.22  | 49.34 | 5.98   |              | 4.52845E-14    |
| X^2   | 3.42  | 0.44  | 0.21  | 24.35 | 12.90 | 12.28  | <b>53.60</b> |                |
| X^2   | 1.16  | 0.15  | 0.07  | 8.22  | 4.36  | 4.15   | <b>18.10</b> |                |
| GCT   | 3.00  | 6.00  | 6.00  | 2.00  | 5.00  | 3.00   | 25.00        |                |
| GCT   | 0.00  | 31.00 | 51.00 | 0.00  | 1.00  | 32.00  | 115.00       |                |
| TOTAL | 3.00  | 37.00 | 57.00 | 2.00  | 6.00  | 35.00  | 140.00       |                |
| Exp.  | 0.54  | 6.61  | 10.18 | 0.36  | 1.07  | 6.25   |              | <b>p-value</b> |
| Exp.  | 2.46  | 30.39 | 46.82 | 1.64  | 4.93  | 28.75  |              | 1.63107E-08*   |
| X^2   | 11.34 | 0.06  | 1.72  | 7.56  | 14.40 | 1.69   | <b>36.76</b> |                |
| X^2   | 2.46  | 0.01  | 0.37  | 1.64  | 3.13  | 0.37   | <b>7.99</b>  |                |
| GCG   | 14.00 | 7.00  | 28.00 | 8.00  | 45.00 | 12.00  | 114.00       |                |
| GCG   | 0.00  | 2.00  | 13.00 | 23.00 | 1.00  | 57.00  | 96.00        |                |
| TOTAL | 14.00 | 9.00  | 41.00 | 31.00 | 46.00 | 69.00  | 210.00       |                |
| Exp.  | 7.60  | 4.89  | 22.26 | 16.83 | 24.97 | 37.46  |              | <b>p-value</b> |
| Exp.  | 6.40  | 4.11  | 18.74 | 14.17 | 21.03 | 31.54  |              | 4.91079E-20    |
| X^2   | 5.39  | 0.91  | 1.48  | 4.63  | 16.06 | 17.30  | <b>45.78</b> |                |
| X^2   | 6.40  | 1.09  | 1.76  | 5.50  | 19.08 | 20.55  | <b>54.37</b> |                |
| GAT   | 12.00 | 6.00  | 13.00 | 7.00  | 54.00 | 8.00   | 100.00       |                |
| GAT   | 0.00  | 42.00 | 0.00  | 0.00  | 0.00  | 42.00  | 84.00        |                |
| TOTAL | 12.00 | 48.00 | 13.00 | 7.00  | 54.00 | 50.00  | 184.00       |                |
| Exp.  | 6.52  | 26.09 | 7.07  | 3.80  | 29.35 | 27.17  |              | <b>p-value</b> |
| Exp.  | 5.48  | 21.91 | 5.93  | 3.20  | 24.65 | 22.83  |              | 1.42774E-27    |
| X^2   | 4.60  | 15.47 | 4.99  | 2.68  | 20.71 | 13.53  | <b>61.98</b> |                |
| X^2   | 5.48  | 18.41 | 5.93  | 3.20  | 24.65 | 16.11  | <b>73.78</b> |                |
| GTA   | 20.00 | 8.00  | 38.00 | 13.00 | 8.00  | 18.00  | 105.00       |                |
| GTA   | 25.00 | 9.00  | 2.00  | 0.00  | 41.00 | 6.00   | 83.00        |                |
| TOTAL | 45.00 | 17.00 | 40.00 | 13.00 | 49.00 | 24.00  | 188.00       |                |
| Exp.  | 25.13 | 9.49  | 22.34 | 7.26  | 27.37 | 13.40  |              | <b>p-value</b> |
| Exp.  | 19.87 | 7.51  | 17.66 | 5.74  | 21.63 | 10.60  |              | 2.86269E-14    |
| X^2   | 1.05  | 0.24  | 10.98 | 4.54  | 13.71 | 1.58   | <b>32.08</b> |                |
| X^2   | 1.33  | 0.30  | 13.89 | 5.74  | 17.34 | 1.99   | <b>40.58</b> |                |
| GTT   | 93.00 | 31.00 | 8.00  | 17.00 | 6.00  | 4.00   | 159.00       |                |
| GTT   | 0.00  | 28.00 | 0.00  | 5.00  | 2.00  | 32.00  | 67.00        |                |
| TOTAL | 93.00 | 59.00 | 8.00  | 22.00 | 8.00  | 36.00  | 226.00       |                |
| Exp.  | 65.43 | 41.51 | 5.63  | 15.48 | 5.63  | 25.33  |              | <b>p-value</b> |
| Exp.  | 27.57 | 17.49 | 2.37  | 6.52  | 2.37  | 10.67  |              | 1.10057E-22    |
| X^2   | 11.62 | 2.66  | 1.00  | 0.15  | 0.02  | 17.96  | <b>33.41</b> |                |
| X^2   | 27.57 | 6.31  | 2.37  | 0.36  | 0.06  | 42.62  | <b>79.29</b> |                |
| GTC   | 46.00 | 15.00 | 14.00 | 7.00  | 6.00  | 2.00   | 90.00        |                |
| GTC   | 10.00 | 0.00  | 0.00  | 0.00  | 34.00 | 66.00  | 110.00       |                |
| TOTAL | 56.00 | 15.00 | 14.00 | 7.00  | 40.00 | 68.00  | 200.00       |                |
| Exp.  | 25.20 | 6.75  | 6.30  | 3.15  | 18.00 | 30.60  |              | <b>p-value</b> |
| Exp.  | 30.80 | 8.25  | 7.70  | 3.85  | 22.00 | 37.40  |              | 3.98886E-28    |
| X^2   | 17.17 | 10.08 | 9.41  | 4.71  | 8.00  | 26.73  | <b>76.10</b> |                |
| X^2   | 14.05 | 8.25  | 7.70  | 3.85  | 6.55  | 21.87  | <b>62.26</b> |                |
| GTG   | 2.00  | 4.00  | 7.00  | 9.00  | 9.00  | 5.00   | 36.00        |                |
| GTG   | 10.00 | 0.00  | 44.00 | 0.00  | 54.00 | 0.00   | 108.00       |                |
| TOTAL | 12.00 | 4.00  | 51.00 | 9.00  | 63.00 | 5.00   | 144.00       |                |
| Exp.  | 3.00  | 1.00  | 12.75 | 2.25  | 15.75 | 1.25   |              | <b>p-value</b> |
| Exp.  | 9.00  | 3.00  | 38.25 | 6.75  | 47.25 | 3.75   |              | 5.25983E-12*   |
| X^2   | 0.33  | 9.00  | 2.59  | 20.25 | 2.89  | 11.25  | <b>46.32</b> |                |
| X^2   | 0.11  | 3.00  | 0.86  | 6.75  | 0.96  | 3.75   | <b>15.44</b> |                |
| GAG   | 32.00 | 13.00 | 74.00 | 6.00  | 18.00 | 15.00  | 158.00       |                |
| GAG   | 8.00  | 20.00 | 0.00  | 36.00 | 0.00  | 36.00  | 100.00       |                |
| TOTAL | 40.00 | 33.00 | 74.00 | 42.00 | 18.00 | 51.00  | 258.00       |                |
| Exp.  | 24.50 | 20.21 | 45.32 | 25.72 | 11.02 | 31.23  |              | <b>p-value</b> |
| Exp.  | 15.50 | 12.79 | 28.68 | 16.28 | 6.98  | 19.77  |              | 1.10443E-26    |
| X^2   | 2.30  | 2.57  | 18.15 | 15.12 | 4.42  | 8.44   | <b>51.00</b> |                |
| X^2   | 3.63  | 4.06  | 28.68 | 23.89 | 6.98  | 13.33  | <b>80.57</b> |                |
| GGA   | 17.00 | 24.00 | 18.00 | 25.00 | 27.00 | 133.00 | 244.00       |                |
| GGA   | 37.00 | 22.00 | 0.00  | 2.00  | 19.00 | 13.00  | 93.00        |                |
| TOTAL | 54.00 | 46.00 | 18.00 | 27.00 | 46.00 | 146.00 | 337.00       |                |
| Exp.  | 39.10 | 33.31 | 13.03 | 19.55 | 33.31 | 105.71 |              | <b>p-value</b> |

|              |       |       |       |       |       |       |        |                |
|--------------|-------|-------|-------|-------|-------|-------|--------|----------------|
| Exp.         | 14.90 | 12.69 | 4.97  | 7.45  | 12.69 | 40.29 |        | 2.37162E-19    |
| X^2          | 12.49 | 2.60  | 1.89  | 1.52  | 1.19  | 7.05  | 26.74  |                |
| X^2          | 32.77 | 6.82  | 4.97  | 3.99  | 3.13  | 18.49 | 70.16  |                |
| GGG          | 4.00  | 9.00  | 9.00  | 9.00  | 61.00 | 67.00 | 159.00 |                |
| GGG          | 1.00  | 1.00  | 0.00  | 0.00  | 0.00  | 18.00 | 20.00  |                |
| <b>TOTAL</b> | 5.00  | 10.00 | 9.00  | 9.00  | 61.00 | 85.00 | 179.00 |                |
| Exp.         | 4.44  | 8.88  | 7.99  | 7.99  | 54.18 | 75.50 |        | <b>p-value</b> |
| Exp.         | 0.56  | 1.12  | 1.01  | 1.01  | 6.82  | 9.50  |        | 0.001994726*   |
| X^2          | 0.04  | 0.00  | 0.13  | 0.13  | 0.86  | 0.96  | 2.11   |                |
| X^2          | 0.35  | 0.01  | 1.01  | 1.01  | 6.82  | 7.61  | 16.80  |                |
| GGC          | 7.00  | 4.00  | 11.00 | 10.00 | 6.00  | 17.00 | 55.00  |                |
| GGC          | 0.00  | 29.00 | 0.00  | 1.00  | 0.00  | 32.00 | 62.00  |                |
| <b>TOTAL</b> | 7.00  | 33.00 | 11.00 | 11.00 | 6.00  | 49.00 | 117.00 |                |
| Exp.         | 3.29  | 15.51 | 5.17  | 5.17  | 2.82  | 23.03 |        | <b>p-value</b> |
| Exp.         | 3.71  | 17.49 | 5.83  | 5.83  | 3.18  | 25.97 |        | 1.52492E-10*   |
| X^2          | 4.18  | 8.54  | 6.57  | 4.51  | 3.58  | 1.58  | 28.97  |                |
| X^2          | 3.71  | 7.58  | 5.83  | 4.00  | 3.18  | 1.40  | 25.70  |                |
| GGT          | 44.00 | 5.00  | 20.00 | 5.00  | 5.00  | 6.00  | 85.00  |                |
| GGT          | 0.00  | 0.00  | 0.00  | 33.00 | 0.00  | 2.00  | 35.00  |                |
| <b>TOTAL</b> | 44.00 | 5.00  | 20.00 | 38.00 | 5.00  | 8.00  | 120.00 |                |
| Exp.         | 31.17 | 3.54  | 14.17 | 26.92 | 3.54  | 5.67  |        | <b>p-value</b> |
| Exp.         | 12.83 | 1.46  | 5.83  | 11.08 | 1.46  | 2.33  |        | 2.92027E-18*   |
| X^2          | 5.28  | 0.60  | 2.40  | 17.85 | 0.60  | 0.02  | 26.75  |                |
| X^2          | 12.83 | 1.46  | 5.83  | 43.34 | 1.46  | 0.05  | 64.97  |                |

\*Kolmogorov-Smirnov test performed and generally consistent with the Chi Square test.

**Table B.** The p-values for the comparisons between dRhodamine and BigDye v1.1 dye chemistries for each of the primers, Primers A1, B1, C1, and D1. Any p-value < 7.8125E-4 (0.05/64) is considered statistically significant with an overall  $\alpha = 0.05$  for that frame given the null hypothesis that the distribution of patterns is the same. All p-values exceeding 7.8125E-4 are black reversed; the null hypothesis is not rejected. Overall, comparisons between dye chemistries yielded statistically significant different distributions of patterns. All N/A correspond to no observations or too few observations for that frame.

| Frame | p-value      |              |              |              |
|-------|--------------|--------------|--------------|--------------|
|       | A1           | B1           | C1           | D1           |
| AAA   | 6.16016E-17  | 2.45218E-07* | 2.45507E-18  | 5.68865E-05  |
| AAC   | 7.42701E-14  | 2.53476E-12* | 1.97785E-15  | 6.67162E-07  |
| ACA   | 1.74519E-05  | 2.65634E-09* | 9.88894E-19  | 2.80704E-27  |
| ACC   | 0.898580318* | 2.53989E-19* | 2.39207E-05  | 8.55863E-17  |
| ACT   | 2.40572E-12  | 9.36644E-47* | 6.01095E-05  | 9.84718E-06  |
| ACG   | 2.39647E-09* | 8.44096E-28* | 5.29652E-11* | 3.47416E-10  |
| AAT   | 8.55696E-15  | 2.99609E-25* | 1.65255E-25  | 8.31074E-06  |
| ATA   | 0.00014045   | 2.02584E-06* | 0.000108357  | 8.39582E-11  |
| ATT   | 9.231E-50    | 2.88039E-28  | 1.02567E-31  | 0.003788378  |
| ATC   | 4.77926E-22  | 1.5425E-31   | 4.86079E-26  | 1.22799E-15  |
| ATG   | 1.42378E-33  | 2.30898E-11  | 8.44338E-33  | 2.49644E-12  |
| AAG   | 1.05149E-39  | 1.26433E-30  | 4.9977E-07   | 2.28566E-26  |
| AGA   | 8.90613E-55* | 5.64318E-17  | 1.54464E-40  | 2.21826E-62  |
| AGG   | 1.1577E-39   | 3.22146E-70  | 8.85743E-31  | 6.58819E-60  |
| AGC   | 9.93676E-38  | 5.9707E-29*  | 1.03736E-39  | 2.32074E-38  |
| AGT   | 2.87085E-28  | 1.31455E-69  | 2.9264E-06   | 1.43778E-12  |
| CAA   | 1.42852E-21  | 1.33925E-09* | 1.05113E-25  | 1.09032E-13  |
| CAC   | 0.495224527* | N/A          | 4.0599E-08   | 1.87137E-08  |
| CCA   | 0.372296673* | 5.75097E-11  | 6.64381E-15  | 2.36207E-09  |
| CCC   | 6.62206E-30* | 1.71613E-12  | 1.40869E-12* | 3.32305E-06* |
| CCT   | 2.0189E-18   | 0.00022521   | 6.30199E-13  | 5.02676E-34  |
| CCG   | 9.40118E-16  | 8.44796E-07* | 1.54692E-08* | 0.047958378* |
| CAT   | 1.37468E-17* | 1.51086E-31  | 2.74849E-05  | 1.4824E-17   |
| CTA   | 1.14762E-08* | 1.35099E-08  | 1.47031E-05  | 3.00249E-05  |
| CTT   | 2.12123E-25  | 2.72968E-30  | 2.6684E-05   | 1.06678E-28  |
| CTC   | 6.88923E-16* | N/A#         | 1.42705E-07  | 4.80945E-26  |
| CTG   | 5.62689E-60* | 7.98395E-39  | 5.34485E-41  | 1.84013E-30  |
| CAG   | 7.88052E-42  | 2.74303E-31  | 3.9521E-43   | 9.19128E-30  |
| CGA   | N/A          | 8.40556E-13* | 2.96881E-13* | 2.78319E-19  |
| CGG   | 5.90774E-12* | 7.48635E-07  | 3.08516E-13* | 1.16316E-15  |
| CGC   | 3.71598E-10* | N/A          | 3.01308E-07  | 1.22487E-07  |
| CGT   | 1.3526E-07   | 6.73872E-13* | 4.96619E-16* | 0.024679084  |
| TAA   | 1.17761E-09  | 3.51214E-27  | 1.53755E-18  | 0.001062979  |
| TAC   | 4.2101E-31   | 1.1553E-09*  | 5.26601E-17* | 1.74571E-11  |
| TCA   | 1.77551E-33* | 0.010295351* | 7.94784E-27  | 6.76191E-13  |
| TCC   | 1.00373E-25  | 4.51852E-45  | 3.05193E-10  | 6.76137E-11  |
| TCT   | 2.41946E-18  | 1.38466E-10  | 2.11567E-08  | 7.92172E-24  |
| TCG   | 2.09432E-14* | N/A          | 3.22366E-11* | 7.06554E-12  |
| TAT   | 7.42205E-23  | 1.15031E-48  | 8.54536E-16  | 1.55971E-27  |
| TTA   | 1.51034E-32  | 1.32654E-21  | 6.21845E-40  | 9.22466E-22  |
| TTT   | 1.27486E-24  | 2.42045E-55  | 1.35895E-47  | 1.66901E-21  |
| TTC   | 1.17765E-14  | 3.16892E-35  | 4.67497E-31  | 2.49665E-09  |
| TTG   | 1.94128E-72  | 4.90151E-16  | 1.16604E-51  | 9.80693E-74  |
| TAG   | 1.62758E-29  | 1.45973E-28* | 1.42781E-17* | 4.78699E-20  |
| TGA   | 1.71049E-65  | 8.46536E-79  | 1.95185E-16* | 1.92539E-09  |
| TGG   | 1.95282E-30  | 1.05512E-50* | 1.93418E-07* | 2.03096E-65  |
| TGC   | 4.64701E-28  | 0.002893949  | 2.01504E-15* | 2.26355E-55  |
| TGT   | 6.72914E-45* | 4.37992E-28  | 1.04523E-65  | 5.08113E-63  |
| GAA   | 0.002901578  | 8.04619E-34  | 1.49219E-22  | 2.7239E-19   |
| GAC   | 1.25675E-14* | 2.00355E-20* | 2.94479E-14* | 1.68173E-27  |
| GCA   | 2.41945E-12  | 1.38709E-19  | 1.58114E-30  | 6.60941E-19  |
| GCC   | 3.33371E-15* | N/A          | 4.52845E-14  | 0.000411963  |
| GCT   | 4.18256E-22* | 3.13126E-13  | 1.63107E-08* | 8.34279E-12  |
| GCG   | N/A          | 3.92678E-10* | 4.91079E-20  | 9.31409E-20  |
| GAT   | 1.77492E-44  | 9.9718E-06   | 1.42774E-27  | 7.42707E-40  |
| GTA   | 5.26794E-08  | 4.02333E-52* | 2.86269E-14  | 1.01273E-09  |
| GTT   | 6.73137E-37* | 2.92847E-07  | 1.10057E-22  | 1.24276E-86  |
| GTC   | 0.002355522  | 9.12813E-20  | 3.98886E-28  | 4.84494E-42  |
| GTG   | N/A          | 1.91843E-11* | 5.25983E-12* | 3.44158E-51  |
| GAG   | N/A          | 9.15203E-41  | 1.10443E-26  | 1.76721E-25  |
| GGA   | 5.01032E-13  | 5.02106E-17  | 2.37162E-19  | 0.008376152  |
| GGG   | 1.05882E-18* | 8.02414E-31  | 0.001994726* | 2.55739E-44  |
| GGC   | N/A          | 1.11532E-15  | 1.52492E-10* | 1.68975E-13  |
| GGT   | 8.69175E-25  | 2.74521E-23  | 2.92027E-18* | 2.45509E-79  |

\*Kolmogorov-Smirnov test performed and generally consistent with the Chi Square test.