Nucleic Acid Tests and Nucleic Acid Amplification Tests for Ginseng Species Authentication Conducted on the Microfluidic Chip

by

Christopher Oberc

Master of Science, Brock University, 2015

Bachelor of Science (Honours), University of Toronto, 2010

Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in the

Department of Chemistry

Faculty of Science

SIMON FRASER UNIVERSITY

Fall 2021

Declaration of Committee

Name: **Christopher Oberc**

Degree: **Doctor of Philosophy**

Title: **Nucleic Acid Tests and Nucleic Acid Amplification Tests for Ginseng Species Authentication Conducted on the Microfluidic Chip**

Committee:

Chair: **Robert Britton**
Professor, Chemistry

Paul C.H. Li
Supervisor
Professor, Chemistry

Dipankar Sen
Committee Member
Professor, Molecular Biology and Biochemistry

Hogan Yu
Committee Member
Professor, Chemistry and Molecular Biology and Biochemistry

Jack Chen
Committee Member
Professor, Molecular Biology and Biochemistry

Bingyun Sun
Examiner
Associate Professor, Chemistry

Hugh Fan
External Examiner
Professor, Biomedical Engineering
University of Florida

Abstract

Panax ginseng and *Panax quinquefolius*, which are commonly called Chinese/Korean ginseng and Canadian/American ginseng, have different medical properties and market values; however, they can be difficult to differentiate from one another based on physical appearances especially when the samples are in powdery or granular forms. A molecular technique is thus needed to overcome this difficulty; this is based on the nucleic acid test (NAT) conducted on the microfluidic chip. Three single nucleotide polymorphism (SNP) sites on the *Panax* genome that differ between. *P. ginseng* and *P. quinquefolius* have been studied using four different NAT methodologies involving probe hybridization to one of the three SNP sites (N1, N2, N3) on the antisense strands amplified by asymmetric PCR. These NATs are distinguished by what is immobilized on the microfluidic chip surface in the first step (i.e. probe, target or capture strand) and by the liquid flow method (i.e. static or dynamic). These methods are probe–target method, target–probe method, capture strand–target–probe method and dynamic probe–target method. Several new probe/target design rules for the NATs have been developed from this study. Out of the four methods, it was found that the capture strand-target-probe method provided the best differentiation, in which a 3'-NH_2 capture strand is first immobilized, the antisense PCR strand is then bound, while N2G and N3Q probes are finally used for detection of *P. ginseng* (G) and *P. quinquefolius* (Q), respectively.

One of the three SNP sites (N2) has been selected to differentiate different *Panax* species using a nucleic acid amplification test (NAAT) known as lesion-induced DNA amplification (LIDA) which has been developed for use in the microfluidic chip. This isothermal NAAT not only amplifies the extracted plant genomic samples but also allows for detection of specific SNPs better than the conventional NAT. This method was used to authenticate 14 ginseng samples (powdery, granular, root); next generation sequencing by Illumina was used to verify the NAAT results.

Keywords: NAT; DNA Hybridization; SNP; Microfluidics; Ginseng Species Authentication; LIDA

Dedication

I would like to dedicate this thesis to my family who have supported me throughout my thesis and helped pull me through the loneliness created by the pandemic.

Acknowledgements

I would like to thank my senior supervisor Dr. Paul C.H. Li for supporting me throughout my PhD program and helped me to think critically about analyzing results. I would also like to thank my other committee members: Dr. Dipankar Sen, Dr. Hogan Yu and Dr. Jack Chen for their contributions and constructive criticism of my work.

I am grateful for Mahsa Gharibi Marzancola who had trained me in performing the hybridization microarray experiments used throughout my thesis work and Abootaleb Sedighi who had provided advice to help me with various aspects of DNA hybridization. In addition, I would like to thank Parwaz Brar for helping me with the centrifugal hybridization flow rate measurements.

I also like to thank my mother who has helped me greatly with script writing required in analyzing the NGS FASTQ files.

Finally, I would like to thank any other faculty members, technicians and students that have contributed towards my degree.

Table of Contents

Declaration of Committee ..... *ii*

Abstract ..... *iii*

Dedication ..... *iv*

Acknowledgements ..... *v*

Table of Contents ..... *vi*

List of Tables ..... *ix*

List of Figures ..... *xi*

List of Acronyms ..... *xvii*

Glossary ..... *xix*

Chapter 1 Introduction *1*

a) Multiplex Nucleic Acid Tests **1**

i) Probe Immobilization 1
ii) Hybridization of Target DNAs 4
iii) Detection of Successful Hybridization 8

b) Hybridization Methodologies **9**

c) Sequencing **10**

i) First Generation Sequencing 10
ii) Second Generation Sequencing 11
iii) Third Generation Sequencing 11

d) Ginseng Species Differentiation **12**

i) Dammarenediol-II Synthase 12
ii) Medicinal Effects of Ginseng 14
iii) Methods of Chemical Authentication of Ginseng 15
iv) Methods of Genetic Authentication of Ginseng 15
v) Research Objectives 16

Chapter 2 Material and Methods *18*

a) Materials **18**

i) Reagents and Solvents 18
ii) Buffers 18

b) Chip Production and Glass Functionalization **18**

i) Si/SU-8 Mold for Chip Fabrication 18
ii) PDMS Slab Casting 19
iii) Aldehyde Glass Functionalization 20

c) Ginseng Probe and Primer Design **21**

i) Sequence Information 21
ii) PCR Primer Sets 21
iii) Genomic DNA Extraction 24
iv) PCR Amplification 25

v) Probe and Target Strands for Detection of Three SNP Sites....25

d. Sample Preparation for DNA Sequencing....29

i) Sanger Sequencing....30

ii) Illumina Sequencing....30

e. Thermodynamics Softwares (DINAMelt, Tm, Mfold)....31

f. Microfluidic Chip Experiment Procedure....32

i) Strand Immobilization....32

a) Immobilization of Oligonucleotide Strands....32

b) PCR strand Immobilization....32

ii) Hybridization....33

a) Stop Flow Hybridization....33

b) Dynamic Centrifugal Hybridization....33

iii) Fluorescence Labeling and Washing....34

iv) Scanning and Image Analysis....34

Chapter 3 Different Strategies for Nucleic Acid Tests Using PCR Amplified Strands..36

a) Immobilized Probe-Target Hybridization (Method 1)....36

i) Utilization of the N1 and N2 Sites....36

ii) Utilizations of Shorter Probes (N1 and N2) and the N3 Site....37

iii) Comparison of Three Lengths of PCR Strands....39

b) Immobilized Target-Probe Hybridization (Method 2)....40

i) Immobilization Time....41

ii) Immobilization of oligonucleotides vs. PCR Strand....42

iii) Comparison of the Lengths of Probes, and N3 SNP Site....45

iv) Long (226bps) vs. Short (126bps) Products....46

c) Immobilized Capture Strand-Target-Probe Hybridization....47

i) Using a 5'-NH_2 Capture Strand....48

ii) Comparing PCR Strands of Different Lengths....49

iii) Using a 3'-NH_2 Capture Strand....50

d) Dynamic Hybridization using Centrifugal Chip Holder....51

i) Probe-Target Hybridization (Method 1) vs. Target-Probe Hybridization (Method 2)....52

ii) Optimization of Probe-Target Method: Intensity Enhancement in Dynamic Hybridization....53

iii) Probe-Target Hybridization of PCR Products (Method 1)....54

e) Comparison of the Three Different Nucleic Acid Test Methods....56

f) Conclusions on the Nucleic Acid Tests of PCR Amplified Strands....58

Chapter 4 On-Chip Nucleic Acid Amplification Test using Unamplified Genomic DNA....60

a) Introduction to SNP Detection Technology....60

b) Materials and Methods....61

c) Optimization Using ds gBlock Templates....69

i) Reaction Time of LIDA (Need a 2,4,6 h Experiment)....70

ii) Free NH_2-Oligo Added to the Ligation Mixture....71

iii) Effect of Different Amounts of [Mg^{2+}]....73

iv) Altering the Amount of QSF and Ligase 74

d) Analysis of 'Known' and Unknown Ginseng Samples 74

i) Pure ds Templates (gBlocks) 74

ii) Gin/Quin Ratios of Unknown Ginseng Samples 77

e) Consideration for an Internal Standard 83

f) Conclusion 87

Chapter 5 Conclusions and Future Work 88

Chapter 6 References 91

Chapter 7 Appendices: 102

Appendix A: Image analysis using Image Lab 102

Appendix B: Melting Temperature Data 111

Appendix C: Raw Fluorescence Microarray Images Produced in Multiplex Nucleic Acid Tests 112

Appendix D: Supporting Figures 133

Appendix E: Sanger Sequencing Results of PCR products 141

Appendix F: Illumina NGS Supporting Data 144

Appendix G: Secondary Structures of 226 nt and 126 nt PCR Products 152

List of Tables

Table 1: List of surface functional groups used as well as their advantages and disadvantages for DNA probe immobilization based on three immobilization techniques.2

Table 2: The nearest-neighbour ΔH, ΔS and ΔG values of each of the 10 different sets of two-nucleotide bp sequences. All values are based on the solution containing 1 M NaCl at 37°C [28]. Copyright permission from ACS. 7

Table 3: Various methods employed for the detection of duplex formation. 8

Table 4: List of all primers and their functionalization. Note that the P6 and P7 primers each have two different functional groups on the same primer sequence (amine for target-probe method; biotin for probe-target method). 22

Table 5: List of all oligonucleotide strands with their respective functionalization, sequence, and purpose in the hybridization experiments. Probes with names containing N1, N2 and N3 hybridize to the SNP sites (N1, N2 and N3, respectively) shown in Figure 2.3. The Probe-Target method is the conventional Method 1 used that involves immobilizing a short probe and hybridizing it with a longer length PCR strand (Figure 2.5a) while the Target-Probe method is the new Method 2 being developed for immobilizing the long PCR product and hybridizing it with a shorter-length probe (Figure 2.5b). Method 3 is the three-strand method that used a capture strand (immobilized) to capture the PCR strand, followed by hybridizing it with the detection probe (Figure 2.5c). 28

Table 6: List of all PCR primers used for the Illumina sequencing. In the PCR1 primers, the bold sections are locus-specific that are the same as the P8' and P7 primers; the italic sections form toeholds on the PCR1 product during the first-round amplification. The complementary sections of these parts of the toeholds (red or blue) hybridize to the 3’ end of the PCR2 primers (red or blue) in the second-round amplification process which forms a set of adapters. These adapters contain the universal primers (italic) and indices (underlined) in the PCR2 primers which are used in bridge amplification in Illumina sequencing. 31

Table 7: Heat map showing relative binding (ΔG) based on DINAMelt thermodynamics of the originally long probes (N1G, N2G, N1Q and N2Q) and short counterparts with the 226 nt PCR strands. The darker the rectangle is in the heat map, the more negative is the ΔG value of hybridization. 44

Table 8: The comparison of the three probes to each of the three methods as well as whether SF or dynamic (Dyn.) hybridization was used. “Yes” indicates that the method was able to successfully differentiate X3 from AmG2 and “No” indicate that the method could not differentiate the two samples. 57

Table 9: List of immobilization and SF hybridization times of each of the three methods used. All hybridization times are reduced (to 10 min) when dynamic hybridization is performed. 57

Table 10*:* Tm calculations used to decide on N2 instead of N1 for designing LIDA oligonucleotides. The table contains a list of all the duplexes that can form among the various DNA fragments and templates for the sequences used by Kausar *et al* at U of A (a),[107,121] by the proposed ginseng N1 site (b) and by the proposed ginseng N2 site (c). "Temp" referred to the *P. ginseng* sense template that helped produce the ligated product of "CompSNP Frag pm" and "CompAbFrag"; "AbCompTemp" refers to this ligated product. "SNP Frag mm" and "CompSNP Frag mm" are similar sequences to "SNP Frag pm" and "CompSNP Frag pm", respectively containing a single mismatching nt each, and "CompTemp" is a perfectly complementary sequence to "Temp". "Ab" in any of the names indicates that the strand contains an abasic position. When the N2 site is used in case c, Temp and CompTemp are the 17 nt sense strand around N2 of *P. ginseng* and *P. quinquefolius*, respectively; CompSNP Frag pm is Gin-NH_2 and CompSNP Frag mm is Quin-NH_2; CompAbFrag is Det; SNP Frag mm is QSF and SNP Frag pm is GSF. Finally, Frag is the same in either the N1 or N2 site (see Table 11). The Tm values highlighted in green pertain to the most relevant duplexes in the amplification reaction. Since DINAMelt interprets the abasic positions as a conventional mismatching bp, the Tms of the "Temp-AbCompTemp" duplexes are overestimated[122] for all three loci. 65

Table 11: List of the oligonucleotides (ss and ds) used for the LIDA reaction. The sense sequence is shown in Figure 4.1 as the template in Cycle 1. The oligonucleotides in this table represents the following strands in Table 10c: Detection Frag (CompAb Frag), Common Fragment (Frag), Gin-NH_2 (ComSNP Frag pm), Quin-NH_2 (ComSNP Frag mm), Gin Specific Fragment (SNP Frag pm), Quin Specific Fragment (SNP Frag mm), ligated Gin Specific Fragment and Common Frag (Temp), and ligated Gin-NH_2 and Detection Frag (AbComTemp). 67

Table 12: The lists of buffer ingredients for T4 ligases obtained from the two vendors NEB and TF. 76

Table 13: Differentiation Gin/Quin ratios calculated from Figure 4.9. 77

Table 14: The ratios (G/Q) of the relative abundance of the P. ginseng and P. quinquefolius genotypes at the N1, N2 and N3 SNP sites of the 14 ginseng samples obtained from NGS through analysis using BWA and IGV. The samples are classified as P. ginseng (G), P. quinquefolius (Q), and heterozygous (H). 80

Table 15: Enthalpies of pairing of Watson-Crick and wobble base pairs.[128] 85

Table 16: Comparison of authentication results of powder and root ginseng samples. ... 89

List of Figures

Figure 1.1: Schematic of immobilizing a DNA probe onto a functionalized solid surface.1

Figure 1.2: Schematic of hybridizing a labeled DNA target to the immobilized DNA probe. 4

Figure 1.3: Schematic of the detection of successful hybridization based on a biotin-labeled duplex that is fluorescently labeled using streptavidin-cyanine 5. 8

Figure 1.4: Hybridization accomplished using multiple surface-immobilized probe lanes and multiple channels with DNA targets passed through. Copyright permission by American Chemical Society.[42] 10

Figure 1.5: The conversion of 2,3-oxidosqualene to dammarenediol-II which is subsequently hydroxylated by ginseng cytochrome P450s and then glycosylated by glycotransferases. 13

Figure 1.6: The common ginsenosides found in P. ginseng and P. quinquefolius where Glc=D-glucose, Ara(p)=L-arabinose (pyranose form), Ara(f)=L-arabinose (furanose form) Xyl=xylose, Rha=L-rhamnose, PPD=protopanaxadiol and PPT=protopanaxatriol. Superscripts indicate the position (refer to image of D-glucose) of the glycosidic linkage.[67–72] 14

Figure 2.1: Image of the 16 channel PDMS slab on a glass slide. The channels have been filled with blue dyed solution for illustration. 20

Figure 2.2: The functionalization of the glass surface with aldehyde groups. a) The hydroxyl surface of the glass is functionalized with amine groups using (3-aminopropyl)triethoxysilane. b) The amine surface is converted to an aldehyde surface using glutaraldehyde. 21

Figure 2.3: A section of the DS gene sequence (sense strand, 5' to 3'). The nucleotides in green represents the five SNP sites; the first character being the one found in P. ginseng and the second being that in P. quinquefolius. The three single SNP sites have been sequentially labeled as N1, N2 and N3, at positions 527, 555 and 586, respectively. The tail of the arrows represents the 5’ end and the head of the arrow represents the 3’end of the primers (blue) and the Int probe (red). The reverse primer sequences are complementary to the sequence presented. 23

Figure 2.4: a) A whole ginseng root (AuChG), b) the central section of a root sample (AmG2) (left) and cross-sectional slice of the root (AmG) (right), c) six ginseng root powder samples (X1-6) and d) AmG3 root sample. 24

Figure 2.5: The orientation of the SNP sites on the PCR strands (antisense) using different techniques a) conventional probe-target Method 1 using an N1 probe (left) and N3 probe (right) b) target-probe Method 2 c) capture-strand-target-probe Method 3..... 27

Figure 2.6: a) Photograph of the centrifugal chip holder with two chips on it. The four letters A, B, C and D represent the angles of the parallel portions of the channels to the radius which are 30°, 60°, 40° and 50°, respectively. Note that for direction B and D, the PDMS slabs have to be sealed perpendicular to the orientation presented in the figure. b) Photograph of the chip holder with two chips on it mounted on the rotating platform. Only one chip is used for an experiment at a time while the other chip functions as a counterweight. 34

Figure 2.7: Image analysis of spots in 16 lanes and data output to a spreadsheet file (same figure as step 14 in Appendix A). 35

Figure 3.1: The probe-target method (conventional Method 1)-immobilized probe hybridized to a PCR strand labeled with biotin. The black, red, blue, and green circles on the PCR strand represent the Int, N1, N2 and N3 hybridization regions, respectively. ... 36

Figure 3.2: The relative intensities of biotin-labelled PCR products (Asymmetric, 226 nt) of the six ginseng powder samples hybridizing to immobilized A-N1G, A-N2G, A-N1Q and A-N2Q (Method 1). The relative intensities have been normalized with Int selected in the consensus region as an internal standard (see Figure 2.3). Raw image data can be found in Appendix C (Figure 7.1). 37

Figure 3.3: Fluorescence intensity of biotin-labeled X3 and AmG3 PCR strands (226 nt) hybridized to the immobilized probes: N2G, N1Q, tN1Q, N3Q, and LN3Q (Method 1). The relative intensities have been normalized with P8'. Raw image data can be found in Appendix C (Figure 7.3). 39

Figure 3.4: Fluorescence intensity of biotin-labeled X3 and AmG2 (126 nt, 226 nt, 694 nt) PCR strands hybridized to the immobilized probes: N2G, tN2G, tN1Q and N3Q (Method 1). This graph is normalized to factor in the differences in MW of the PCR strands. The inset shows the results in the expanded scale for the 694 nt PCR strand. Raw image data can be found in Appendix C (Figure 7.3) 40

Figure 3.5: The target-probe method (Method 2)-immobilized PCR strand hybridized to a detector probe labeled with biotin. The black, red, blue, and green circles on the PCR strand represent the Int, N1, N2 and N3 hybridization regions, respectively. 41

Figure 3.6: Immobilization of 226 nt AmG2 PCR strand to the solid surface using $NaCNBH_3$ and incubating it for 2, 4 and 6 hrs.: B-N2Q was used for detection. Raw image data can be found in Appendix C (Figure 7.4). 42

Figure 3.7: Hybridization of immobilized A-Gin and A-Quin to B-N2G and B-N1Q (Method 2), a) SF hybridization, b) dynamic hybridization (1 injection of 1.0 μL, 5 spins). Raw image data can be found in Appendix C (Figure 7.5). 43

Figure 3.8: Mfold-predicted intramolecular folding of the 226 nt antisense strands of a) P. ginseng and b) P. quinquefolius. Both structures were predicted using 0.15 M Na^+ and 22°C 44

Figure 3.9: Hybridizations of immobilized 226 nt X3 and AmG2 PCR strands to B-N1G, B-N2G, B-N1Q and B-N2Q detection probes (Method 2): a) SF hybridization (1 h) b) Centrifugal (dynamic) hybridization using a single 1.0 μL injection. Raw image data can be found in Appendix C (Figure 7.5). 45

Figure 3.10: SF hybridization of immobilized 226 nt X3 and AmG2 to B-N2G, B-tN2G, B-N1Q, B-tN1Q and B-N3Q (Method 2). The hybridization was performed using the 1 h SF method. Raw image data can be found in Appendix C (Figure 7.6). 46

Figure 3.11: Comparison of the effect of the length of PCR strands on their hybridization to the probes B-N2G, B-tN2G, B-N1Q, B-tN1Q and B-N3Q (Method 2). All hybridization was done using the 1 hr SF method. a) X3 and X3' are the 226 nt and 126 nt PCR strands, respectively. b) AmG2 and AmG2' are the 226 nt and 126 nt PCR strands, respectively. The intensities of the 226 nt PCR strands in a) and b) were used to construct the graph in Figure 3.10. Raw image data can be found in Appendix C (Figure 7.6). 47

Figure 3.12: Schematic of a three-strand complex using a 5' amine capture strand, unlabeled PCR strand and a detector probe labeled with biotin (Method 3). 48

Figure 3.13: Three strand complexes made by hybridizing unlabeled 226 nt asymmetric X3 and AmG2 to immobilized 5'-A-P8' capture strand (Method 3). Then B-N2G, B-tN2G, B-N1Q, B-tN1Q, B-N3Q are hybridized to those captured PCR strands (SF, 1 h). Each three-strand complex was done in six replicates. Raw image data can be found in Appendix C (Figure 7.7). 48

Figure 3.14: Three strand complexes made by hybridizing unlabeled X3 and AmG2 (40ng/μL) to immobilized 5'-A-P8' capture strand (Method 3). Then B-N2G, and B-N1Q are hybridized to those captured PCR products. Each three-strand complex was done in six replicates and both hybridization steps were done using SF hybridization. The top graph was constructed using the fluorescence intensities and the bottom graph was constructed from the quotient of those same intensities over the strand's concentration in nanomolar. Raw data can be found in Appendix C (Figure 7.8). 49

Figure 3.15: Schematic of a three-strand complex using a 3' amine capture strand (primer), unlabeled PCR strand and a detector probe labeled with biotin (Method 3)..... 50

Figure 3.16: Three-strand complexes made by hybridizing unlabeled 226 nt X3 and AmG2 to immobilized 3'-A-P8' capture strand as the capture strand (Method 3). Then B-N2G, B-tN2G, B-N1Q, B-tN1Q, B-N3Q are hybridized to those captured PCR strands. (a) Both hybridization steps were performed using SF hybridization b) The PCR strand hybridization was performed using SF hybridization and the probe hybridization was performed using dynamic hybridization. All three-strand complex were made in six replicates. Raw image data can be found in Appendix C (Figure 7.7). 50

Figure 3.17: Section of a chip image where the solutions containing the amine functionalized oligonucleotides were applied to the horizontal channels and the

hybridizing solution flowed through the vertical channels. The channels on the left half (-8 to -1) have the fluid flow from the bottom to the top and the channels on the right half (1 to 8) have the fluid flow from the top to the bottom. Fluid flow direction is also indicated by arrows in the vertical channels. 51

Figure 3.18: Hybridization of Gin/Quin targets to the N2G and N1Q probes using 2 injections of 0.5 μL. The channel numbering refers to the numbering used in Figure 3.17. Each hybridization was made in quadruplicates. a) hybridization of immobilized targets to biotin labeled probes (Method 2). b) Hybridization of immobilized probes to biotin labeled targets (Method 1). Raw image data can be found in Appendix C (Figure 7.9). . 53

Figure 3.19: Comparison of the target-probe Method 2 using one-spin, target-probe Method 2 using five-spins and the probe-target Method 1 using five-spins. The signals for the probe-target Method 1 using one-spin were so low that they were not shown. Raw image data can be found in Appendix C (Figure 7.10). 54

Figure 3.20: Centrifugal probe-target Method 1 hybridizing asymmetric biotin labeled 226 nt X3 and AmG2 to immobilized N2G and N1Q. The five-spin method was used with two injections of 0.5 μL each. Each injection was spun five times and the solutions was allowed to backflow after each spin. The hybridization buffer contained a) no additives, b) 2.5% formamide (v/v), c) 5.0% formamide (v/v) or d) 10% formamide (v/v). Raw image data can be found in Appendix C (Figure 7.11). 55

Figure 4.1: LIDA with two cycles in the ligation chain reaction. In Cycle 1, one template (A, GinT, sense) hybridizes (step 1.1) to the immobilized site-specific probe (Gin-NH_2, antisense) and causes ligation (step 2.1) of the detection fragment (Det, antisense). The duplex subsequently dissociates (step 3) to regenerate the template for Cycle 1 to repeat. Then, in Cycle 2, the immobilized long strand hybridizes (step 1.2) to site-specific fragment (GSF, sense) and causes its ligation (step 2.2) with the common fragment (Frag, sense). The duplex subsequently dissociates (step 3) to give an additional template (B, which is shorter than A) to enter Cycle 1. Therefore, Cycle 1 produces immobilized Det (fluorescence detection) to as many as the limiting amount of Det in the solution; Cycle 2 adds more templates (for Cycle 1) to as many as the amount of GSF or Frag, whichever is less..... 69

Figure 4.2: Effects of reaction time on LIDA. Three reaction times were used for the ligation reaction with Quin-NH_2 and Quin gBlock ds samples as templates using QSF. All the data bars are the averages of quadruplicates, and the error bars are the standard deviations. NEB ligase was used. Raw image data can be found in Appendix C (Figure 4.2). 70

Figure 4.3: The effect of adding free-solution Gin-NH_2 or Quin-NH_2 in the ligation mixture. Gin-NH_2 or Quin-NH_2 (10 nM) were added to the ligation mixture, to be compared with controls containing no free-solution Gin-NH_2 or Quin-NH_2. The mixture was incubated for (a) 2 h, (b) 4 h. AmG2 genomic templates and 10 mM Mg^{2+} were used for all ligation reactions. Raw image data can be found in Appendix C (Figure 7.14). ... 71

Figure 4.4: LIDA of gBlock templates. 100 CEU/µL NEB ligase was used and incubation time was 4 h. 10 nM Gin-NH_2 or Quin-NH_2, and 25 µM of GSF/QSF added. Raw image data can be found in Appendix C (Figure 7.15)..... 73

Figure 4.5: Effect of [Mg^{2+}] on LIDA. The ligation buffer already contains 10 mM Mg^{2+} and so additional Mg^{2+} was added to the buffer to result in the final concentrations of 15, 20, 25 and 30 mM. The blue bars are measured using the left vertical axis and the orange bars are measured using the right vertical axis. Raw image data can be found in Appendix C (Figure 7.16)..... 73

Figure 4.6: The effect of [QSF] on gQuin signals on Quin-NH_2 lanes. gQuin (1.0 pg/µL) was used as the template; Quin-NH_2 was added; 100 CEU/µL ligase was used with 4 h incubation. The last bar pertains to 25µM GSF used with 1.0 pg/µL gGin on Gin-NH_2 lanes. Raw image data can be found in Appendix C (Figure 7.17). 74

Figure 4.7: Intensity for 100% gGin, 100% gQuin, and a 50% mix of each (gGin/Quin). The LIDA mixture was composed of 1 µL 2× Oligo/buffer solution (Det, Frag (50 µM each), and 60 µM GSF or QSF, Gin-NH_2 or Quin-NH_2 (20 nM), 2× buffer, 20 mM $MgCl_2$), 0.5 µL template and 0.5 µL ligase (2 µL final volume). The ligation reaction was 4 h. a) 100 CEU/µL NEB ligase b) 1.25 U/µL TF ligase. Raw data can be found in Appendix C (Figure 7.18)..... 75

Figure 4.8: Intensity for 100% gGin, 100% gQuin, and a 50% mix of each (gGin/Quin). The LIDA mixture was composed of 1 µL 2× Oligo/buffer solution (Det, Frag (50 µM each), and 60 µM GSF or QSF, Gin-NH_2 or Quin-NH_2 (20 nM), 2× buffer, 20 mM $MgCl_2$), 0.5 µL template and 0.5 µL TF ligase (2 µL final volume) and was incubated for 2 h. Raw data can be found in Appendix C (Figure 7.19). 76

Figure 4.9: Differentiation of ginseng samples X1 to X6. Genomic templates used; 10 nM Gin-NH_2 or Quin-NH_2 added, 30 µM GSF or QSF, 25 µM Det, 25 µM Frag, 10 mM [Mg^{2+}] (added), 1.25 U/µL TF ligase, 2 h incubation. Raw data can be found in Appendix C (Figure 7.20)..... 77

Figure 4.10: LIDA of the eight ginseng root samples with LIDA. Though AmG, AmG2 and AmG3 were claimed to be American ginseng, only AmG2 and AmG3 can live up to the claims. While AuChG, AuChG2, KorG, and KorG2 and were expected to be Chinese ginseng, only AuChG2 and ChG fit better to the claimed identities. The reaction was carried out using 1.25 U/µL TF ligase and 2 h incubation. Raw data can be found in Appendix C (Figure 7.20)..... 78

Figure 4.11: IGV image of the ChG (top) and KorG (bottom) sequences where the colours represent C (blue), T (red), G (orange), A (green), deletion (white) and insertion (purple). The row of coloured pixels at the top represents the P. ginseng reference sequence. The bars below the reference sequence represent the cumulation of all nucleotides at the specific position; any position that deviates by ≥20% relative to the reference will show up coloured in. Below the bars are the sequences themselves listed on

order of decreasing length. Any position that was coloured in the bars is also coloured in within the sequences. The N1, N2 and N3 sites are labeled above the reference sequence. The ChG image contain 620 sequences and the KorG image contains 633 sequences. ... 79

Figure 4.12: NGS results of the N2 SNP site for AmG, AmG2, AmG3, AuChG, AuChG2, ChG, KorG, and KorG2 using BWA/IGV (top) and the Java program (bottom) where P. ginseng contains a C base, and P. quinquefolius contains a T base. All intensities have been normalized to the total number of sequences for each sample. Raw data for these root samples and for X1-X6 (powder samples) can be found in Appendix F, Table 20, Figure 7.28. 81

Figure 4.13: NGS results of the N1, N2 and N3 SNP sites of X1-X6 (top) and AmG, AmG2, AmG3, AuChG, AuChG2, ChG, KorG, and KorG2 (bottom). P. ginseng contains the TCG genotype and P. quinquefolius contains the CTA genotype. The new genotype is indicated as CCG. All intensities have been normalized to the total number of sequences for each sample. Raw data can be found in Appendix F (Table 22). 83

Figure 4.14: a) Fluorescence intensities at the immobilized fragment InStd-NH_2 used in combination with LIDA solutions containing 30 µM GSF or QSF, 25 µM Frag, 25 µM Det and 1.0 pg/µL gBlock templates. Raw image data can be found in Appendix C (Figure 7.21, Figure 4.14). b) Watson-Crick G-C and A-T base pairs (top) and wobble I-C and I-T base pairs (bottom). Ribose on the nucleotides is indicated by "rib". c) Numbering adopted for purines (A, G) and pyrimidines (C, T). 86

List of Acronyms

APEX	Arrayed Primer Extension
APTES	(3-Aminopropyl)triethoxysilane
AuNP	Gold Nanoparticle
bp	Base Pair
BWA	Burrows-Wheeler Aligner
ddNTP	Dideoxynucleotide Triphosphate
DNA	Deoxyribonucleic Acid
dNTP	Deoxynucleotide Triphosphate
DS	Dammarenediol-II Synthase
EDC	N-(3-Dimethylaminopropyl)-N′-Ethylcarbodiimide
FRET	Fluorescence Resonance Energy Transfer
GUMA	Gold Nanoparticle-Based Universal Oligonucleotide Microarray Assay
HDA	Helicase Dependent Amplification
IGV	Integrated Genomic Viewer
ITS	Intron Targeting SNP
LAMP	Loop-Mediated Isothermal Amplification
LIDA	Lesion-Induced DNA Amplification
NAAT	Nucleic Acid Amplification Test
NAT	Nucleic Acid Test
NGS	Next Generation Sequencing
NHS	N-Hydroxysuccinimide
nt	Nucleotide
PBS	Phosphate Buffer Saline
PCR	Polymerase Chain Reaction
PDMS	Polydimethylsiloxane
PPD	Protopanaxadiol

PPT	Protopanaxatriol
RCA	Rolling Circle Amplification
rt	Room Temperature
rt-PCR	Real Time PCR
SA-Cy5	Streptavidin-Cyanine 5
SDA	Strand Displacement Amplification
SDS	Sodium Dodecyl Sulfate
SF	Stop Flow
SNP	Single Nucleotide Polymorphism
SSC	Saline Sodium Citrate
TF	Thermo Fisher

Glossary

5-Spin Method	A technique used in the centrifugal dynamic hybridization method where the hybridizing solution is allowed to backflow after centrifugation. The process of centrifugation and backflow is repeated four more times (five centrifugations total) before the solution is removed from the channels.
Asymmetric PCR	A PCR where one primer is in excess resulting in the corresponding PCR product strand being synthesized in excess.
Dynamic Hybridization	The process in which the hybridizing DNA strand solution flows continuously through the channels for the duration of the hybridization reaction.
gBlock	A synthetically made ds DNA sequence made through the ligation of smaller synthetic DNA sequences. The ds sequence is subsequently cloned on a plasmid.
gGin	A gBlock containing the *P. ginseng* genotype.
gQuin	A gBlock containing the *P. quinquefolius* genotype.
Hybrid	An individual that is created by crossing two different species.
Lesion-Induced DNA Amplification (LIDA)	An isothermal ligation-based amplification reaction capable of SNP differentiation. The reaction uses an abasic site placed at the ligation site to allow for facile denaturation of the ligated products.
Melting Temperature (Tm)	The temperature at which half of the DNA duplexes have denatured.
Panax ginseng	The species of ginseng commonly known of Chinese ginseng.
Panax quinquefolius	The species of ginseng commonly known as American ginseng.
Stop Flow (SF) Hybridization	The process in which the hybridizing DNA strand solution remain static for the duration of the hybridization reaction.

Chapter 1 Introduction

This thesis entitled " Nucleic Acid Tests and Nucleic Acid Amplification Tests for Ginseng Species Authentication Conducted on the Microfluidic Chip" deal with various nucleic acid tests conducted in a multiplex manner for differentiating between two ginseng species. This introduction covers the various concepts of nucleic acid test, DNA hybridization, DNA sequencing and ginseng species differentiation. The research objectives are also described in detail.

a) Multiplex Nucleic Acid Tests

Multiplex nucleic acid tests (NATs) detect in a sample the presence of nucleic acids, usually with a specific sequence to determine the natures, origins, variations of biological species. The multiplex tests utilize microfluidic devices in order to create an organized arrangement, or a microarray, of deoxyribonucleic acid (DNA) duplexes on a solid surface for biochemical binding assays. There are three main steps involved in performing these tests: immobilization of probes, binding or hybridization of target, and detection of the duplexes. In conventional tests, probes are DNA molecules spotted on a solid surface, and binding the targets to the multiple probes is based on hybridizations. However, in unconventional tests, a dynamic flow has been used for probe immobilization, and ligation or extension is involved in target binding.

Each of these three steps in multiplex nucleic acid tests will be discussed in detail in the subsequent sections.

i) Probe Immobilization

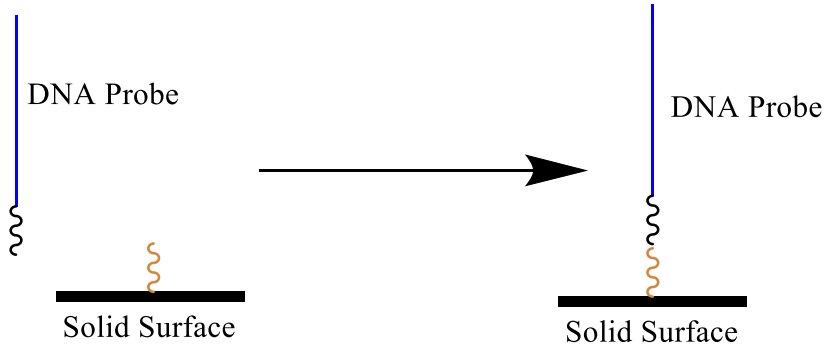


Figure 1.1: Schematic of immobilizing a DNA probe onto a functionalized solid surface.

The first step of multiplex nucleic acid test is to immobilize a probe onto a solid surface (Figure 1.1). There is a variety of different immobilization techniques available for the probes, namely physical absorption, chemical linkage, and biochemical interactions (Table 1).

Table 1: List of surface functional groups used as well as their advantages and disadvantages for DNA probe immobilization based on three immobilization techniques.

Surface Functionalization	**Probe Conjugation**	**Coupling Formed in Immobilization**	**Advantages**	**Disadvantages**	**Ref**
Amine	None	None; Physical Absorption	No label required; Works well at low probe conc.; Reusable surface	Random orientation; loss of immobilized probe after washing	[1]
Poly-lysine	None	None; Physical Absorption	No label required; Amount immobilized proportional to conc.; Reusable surface	Random orientation; loss of immobilized probe after washing	[1,2]
Carboxylic Acid	Amine	Amide	Can be done with nanomolar conc. of probe; Fast coupling reaction	Requires long activation time	[3]
Aldehyde	Amine	Schiff base to be reduced to a secondary amine	Reproducible spotting; Fast reaction	Requires micromolar conc.; Requires reduction step	[4,5]

Epoxide	Amine	Secondary Amine	Robust linkage formed	Slow reaction; Spotting not reproducible	[6]
Isothio-cyanate	Amine	Thiourea	Fast reaction; Less labile to Tris buffer	coupling deteriorates after multiple denaturation; Spotting not reproducible	[7]
Gold	Thiol	Gold-Thiol Coordination	Fast reaction; Reproducible spotting	Labile in acidic conditions	[8,9]
Biotin	Streptavidin	Streptavidin-Biotin Interaction	Good immobilization efficiency	Cannot be used at high temperature or denaturing conditions	[10]
Streptavidin	Biotin	Streptavidin-Biotin Interaction	No dependence on ionic strength Works with nanomolar conc. Uniform immobilization	Cannot be used at high temperature or denaturing conditions	[10,11]

Out of the several immobilization techniques presented in Table 1, the three most common methods are aldehyde-amine coupling, gold-thiol interaction and the streptavidin-biotin interaction. These three methods all have good reproducibility and involve relatively fast coupling reactions. While the coupling reaction between the aldehyde-functionalized surface and the amine-conjugated probe is spontaneous, this reaction that produces the Schiff base is reversible. In order to prevent any dissociation in subsequent steps, the Schiff base formed is converted, using a reducing reagent (usually $NaBH_4$), to a secondary amine which is no longer labile.[5]

Amide immobilization due to amine-carboxylic acid coupling is also used frequently in DNA immobilization because of its fast coupling, however, it is not used as frequently as the other three methods because this coupling is not a spontaneous reaction with a long activation time of the carboxylic acids. The activation must be carried out with N-(3-dimethylaminopropyl)-N′-ethylcarbodiimide (EDC) and N-hydroxysuccinimide (NHS), and the stability of the NHS esters that are formed is low.[3]

For surfaces functionalized with epoxide groups, it has been shown that when the surface is also coated with tertiary amines, the reaction rate and uniformity of the immobilized DNA probes is greatly improved. The explanation for this is that the tertiary amine (pKa~12) on the surface, while it will not react with the epoxides on the surface, brings the DNA probe close to the surface via electrostatic interactions. The amine-conjugated DNA probe now being close to the surface allows it to more easily react with the epoxides on the surface via its primary amine.[1]

In this thesis, the amine-aldehyde coupling method will be used for immobilizing the DNA strands.

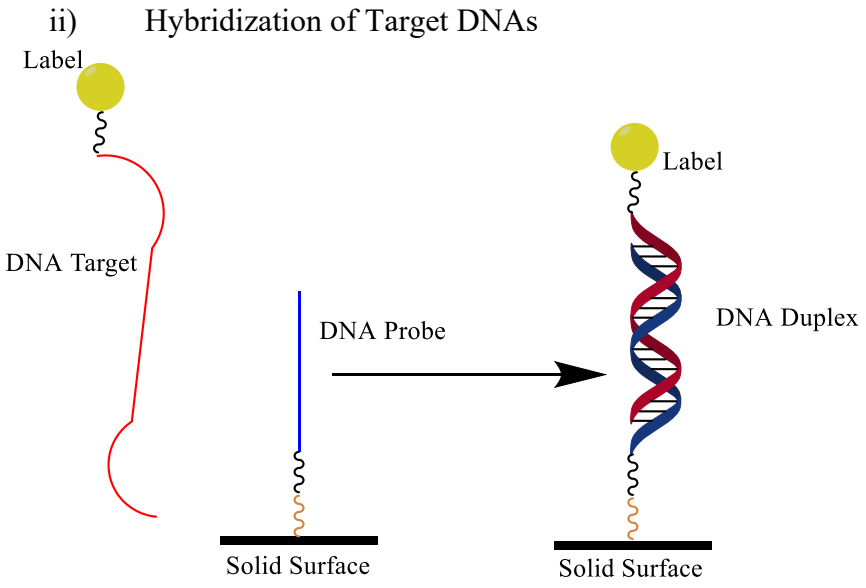


Figure 1.2: Schematic of hybridizing a labeled DNA target to the immobilized DNA probe.

The second step of nucleic acid tests involves capturing a target strand for detection (Figure 1.2). The conventional method is by hybridizing a target strand to the DNA probes which have been immobilized in the first step. This process works on the principle that since the immobilized probe and target have complementary regions in their sequences, they will form a duplex thus capturing the target strand for detection. If the sequences of the two strands are not complementary, the target is washed away and thus not detected.

Usually, the hybridization step is conducted in buffers of a high ionic strength in order to minimize strand repulsion and to maximize the contact, and hence, the hybridization of the target strand to the probe.[12]

A common objective of conducting DNA hybridization experiments is to confirm single nucleotide polymorphisms (SNPs) in a specific region of the DNA sequence. While the interaction between a perfectly matched base pair (bp) is strong, the presence of a mismatched bp weakens the interaction between the two strands. There are various methods used to prevent or lessen the hybridization of the 1-bp mismatched targets. One common method is to perform a wash using a buffer of a low ionic strength as this will cause the negatively-charged sugar-phosphate backbones to repel each other to a greater degree compared to a high ionic strength buffer, thus weakening and denaturing the mismatched duplex.[13,14]

A second method for preventing strands with a mismatched SNP site from hybridizing is to use elevated temperatures during the hybridization. Like the scenario with the low ionic strength buffer, elevated temperatures will denature the less stable mismatched duplexes to a greater extent than perfectly complementary duplexes.[13,15]

A third method is to use a soft denaturing agent (e.g. formamide[16], urea[17], gold nanoparticles[18,19]). Formamide and urea are small planar molecules that will intercalate into the mismatched duplex and weaken it by disruption of the π-π stacking between the nitrogenous bases.[20] Similarly, the gold nanoparticles (AuNPs) will compete with the normal base pairing by coordinating with the bases.[19] This AuNP methodology, which was initially developed by Wang *et al.* requiring the AuNPs to be mixed with the hybridizing strands,[21] was subsequently improved by Sedighi *et al.* by introducing the nanoparticles in a post-hybridization wash buffer.[18] Graphene oxide has also been reported for use in differentiation of SNPs. In contrary to perfectly matched duplexes, the planar graphene regions intercalate into the mismatched DNA duplex and disrupt the π-π stacking of the bases, and the oxide functional groups compete with the H-bonding between the base pairs.[18,22]

Many review articles have been published discussing different aspects of DNA hybridization on solid surfaces.[23–27]

In this thesis, a high ionic strength buffer will be used for the hybridization of the DNA strands and formamide and urea will be used in the buffer in a few circumstances.

Successful multiplex nucleic acid tests by DNA hybridizations depend very much on the proper designs on probes and primers. DNA hybridization/melting software is an important tool for designing probes and polymerase chain reaction (PCR) primers since the tool allows one to examine the thermodynamics of hybridization between two DNA strands without having to synthesize the strands and perform experiments to test them. There are many DNA hybridization algorithms or softwares available;[28–31] which use the nearest neighbour theory to calculate the thermodynamic parameters (e.g. ΔH, ΔS, ΔG) of hybridizations of duplexes.[32,33] The nearest neighbour theory works by summing the ΔH

and ΔS values (at 37°C) for each set of two consecutive complementary bps in the duplex. Even though there are 4^2 or 16 possible combinations for a two-nucleotide bp sequence, only 10 of them are unique sequences, see Table 2. For example, in the three-nucleotide bp sequence CTG/GAC (5'-3'/3'-5'), the first and second bp are CT/GA which have a ΔG of -1.16 kcal/mol (Table 2 entry #6). The next set of two bps is TG/AC (second and third bp) which have a ΔG of -1.38 kcal/mol (Table 2 entry #4, note that the sequence of CA/GT is inverted without loss of generality). Thus, the three-nucleotide bp sequence has a ΔG of hybridization of -(1.16 +1.38) = -2.54 kcal/mol.[34] The value of –ΔG is to be maximized in order to decide on which sequence will result in the most stable hybridization with less intramolecular folding. These algorithms also have an addition function which factor in the reaction conditions such as the ionic strength and temperature.[35]

Table 2: The nearest-neighbour ΔH, ΔS and ΔG values of each of the 10 different sets of two-nucleotide bp sequences. All values are based on the solution containing 1 M NaCl at 37°C [28]. Copyright permission from ACS.

	Two-nucleotide bp Sequence	ΔH° (kcal/mol)	ΔS° (cal/mol·K))	ΔG° (kcal/mol)
1	AA/TT	-8.4	-23.6	-1.02
2	AT/TA	-6.5	-18.8	-0.76
3	TA/AT	-6.3	-18.5	-0.60
4	CA/GT	-7.4	-19.3	-1.38
5	GT/CA	-8.6	-23.0	-1.43
6	CT/GA	-6.1	-16.1	-1.16
7	GA/CT	-7.7	-20.3	-1.46
8	CG/GC	-10.1	-25.5	-2.09
9	GC/CG	-11.1	-28.4	-2.28
10	GG/CC	-6.7	-15.6	-1.77

The nearest neighbour theory only predicts the thermodynamics of hybridization and does not take the kinetics of hybridization into consideration. However, the software for predicting the kinetics of hybridization has been developed which predicts the rate constant (k_{hyb}) of the pairing of sequences, at a certain temperature, and ionic strength (nablab.rice.edu/nabtools/kinetics). The strands that have strong intramolecular folding and/or repeats of three or more bases will have slower hybridization rates than the sequences without these problematic features.[36]

The nearest neighbour theory for predicting the thermodynamics of DNA hybridization *in silico* has been well established for solution-phase hybridization. The DINAMelt and Mfold software[29] will be used in this thesis in designing the probes and primers, both in terms of having an appropriate length (and thus Tm) as well as checking for problematic intramolecular folding.

While *in silico* studies have been reported on DNA hybridization occurring on a solid surface,[37] including in microfluidic settings,[32] there is no software yet available for predicting the thermodynamics or kinetics of such hybridization reactions.

iii) Detection of Successful Hybridization

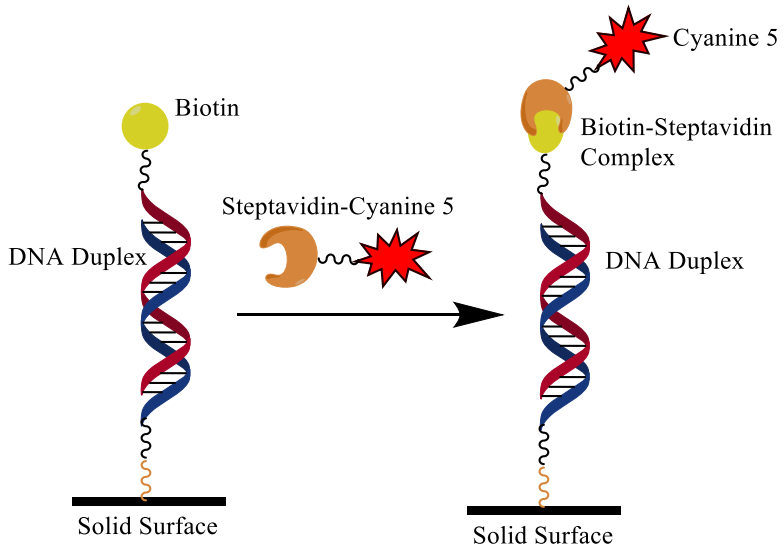


Figure 1.3: Schematic of the detection of successful hybridization based on a biotin-labeled duplex that is fluorescently labeled using streptavidin-cyanine 5.

The final step of nucleic acid tests is the detection of the duplex formation. There are many detection methods, tabulated in Table 4. The most common means of detection is the use of fluorescence. Typically, the target DNA strand is conjugated directly with a fluorophore label prior to being hybridized to the probe (e.g. conjugated using labeled primers in DNA amplification such as PCR).[13,22] Fluorescence detection of the target strand can also be achieved by a fluorophore cyanine 5 (Cy5). This is implemented by conjugating the target strand with a biotin label in the first step and then coupling it with streptavidin-cyanine 5 (SA-Cy5) in the second step (see Figure 1.3).[12]

Table 3: Various methods employed for the detection of duplex formation.

Detection methods	**Label-Duplex Association**	**Reference**
Fluorescence	Direct labeling of target DNA	13,38
Fluorescence	Indirect labeling of target (via biotin conjugation and then coupled to a streptavidin-label)	12
Fluorescence	Base pairing between target cytosine and a porphyrin analogue	39
Silver Stain (Visual)	Staining of AuNP-conjugated target with silver	40

Another technique for SNP differentiation that does not require any conjugation of the target was developed by Lin *et al*. In this technique there is an abasic site in the probe at the position of the SNP site. For detection, a porphyrin analogue with phenolic groups is used which will insert into the abasic site and pair up with the target SNP base. If that

base is cytosine (C), then the porphyrin complex becomes fluorescent and if it is any other base (guanine, thymine, or adenine), then it remains non-fluorescent. This works if the sense strand has a C genotype, but if the sense strand has a G genotype, the antisense strand (C) should be used for hybridization. This presents a limitation to this method since if neither genotype has a G-C base pair (i.e. A-T vs. T-A), the method does not work because there would never be any fluorescence generated.[39]

One method that is not based on photometric detection is the visual detection by the unaided eye using the silver stain. After the hybridization of an AuNP-conjugated target DNA strand to an immobilized probe, an $AgNO_3$/hydroquinone solution is added. This results in a reduction reaction forming metallic silver on the AuNP, which will visually appear as a black/purple stain. This method is not frequently used because it is very difficult to have quantifiable and reproducible results.[40] Photometric methods (e.g. fluorescence and chemiluminescence) are much more robust and reproducible, thus making them popular detection methods in nucleic acid tests.

In this thesis, the indirect fluorescence labeling method is adopted, which involves using biotin-labeled DNA strands and forming a fluorescent species using SA-Cy5.

b) Hybridization Methodologies

There is a variety of different designs of multiplex nucleic acid tests available. Conventionally, these tests are performed either by immobilizing probes on a microwell plate (e.g. 96 well plate),[15] or spotting them onto a functionalized microchip surface (usually glass).[41]

There are, however, unconventional ways to do hybridization-based multiplex nucleic acid tests. One such example is by first applying the probes by spotting in a microwell or on a microchip surface, and then using dynamic flow, instead of static or stop flow (SF), during the hybridization step. This is performed using a microchannel chip made from polydimethylsiloxane (PDMS) that allows a small volume of target solution (e.g. 1 μL) to shuttle back and forth over an array of spotted probes.[12] Another way is by first applying the probes by flowing them in microfluidic channels, followed by the hybridization step. This system has been reported by Wang *et al.* consisting of an aldehyde functionalized glass slide sealed with a PDMS slab containing 16 microchannels. First, the probe solutions are allowed to flow into the microchannels, then the PDMS slab is removed and the reduction step by $NaBH_4$ is performed to complete the surface immobilization of probes. Afterwards, the PDMS is reapplied perpendicular to the original probe lane direction and the targets solutions are applied to the channels (either in a dynamic or static/stop flow fashion). Thus, DNA duplexes only form at the intersections where complementary probes and targets meet (Figure 1.4).[42]

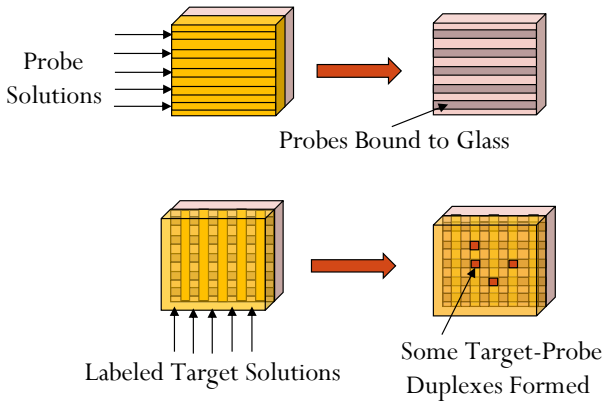


Figure 1.4: Hybridization accomplished using multiple surface-immobilized probe lanes and multiple channels with DNA targets passed through. Copyright permission by American Chemical Society.[42]

This is the hybridization methodology that is adopted for this thesis work. The aldehyde-amine coupling is used for probe immobilization since the reaction is fast and the probe printing is reproducible. Even though a reduction step is required after the coupling, it is also a fast reaction thus not creating much of a time constraint. The target strands are conjugated with biotin and hybridized to the immobilized probes using a high salt buffer. In some cases, formamide or urea is added to the buffer. The biotin then forms a complex with SA-Cy5 which makes the duplex fluorescently labeled. The reason why the biotin-streptavidin coupling is used for fluorescent labeling the oligonucleotides rather than having fluorescent label directly on them is because it is less expensive to purchase biotin conjugated oligonucleotides as well as SA-Cy5 than it is to purchase fluorescently labeled oligonucleotides.

c) Sequencing

Sequencing is the process of deciphering the order of nucleotides of a section of DNA. This process can be used to detect SNPs as well as to detect insertions and deletions in the sequence. There are different sequencing techniques classified by several generations.[43]

i) First Generation Sequencing

The first-generation sequencing is more commonly known as Sanger sequencing which requires the sample sequence to be pure by cloning. It was developed in 1977 using four parallel PCR, each reaction containing one of the four dideoxynucleotide triphosphates (ddNTPs) that terminate the product elongation prematurely to produce fragments of different lengths. The fragments of these reactions were subsequently separated by gel electrophoresis to determine the length and sequence of the fragments which were subsequently assembled to generate the sequence of the sample.[44] This method has been further developed over the years to allow for all four ddNTPs to be used in a single reaction by labeling ddCTP, ddATP, ddGTP and ddTTP each with a different fluorophore and by analyzing the fragments on capillary electrophoresis instead of a traditional gel.[45] The limitation of this technique is that it can only sequence fragments that are <1000

bases, and thus is not an efficient technique for sequencing large genomes. This is why the human genome project took a decade to complete.[46]

A variant of Sanger sequencing exists, and it is designed to identify a SNP site at a specific position and is called arrayed primer extension (APEX). This technique works by designing primers that hybridize at the base immediately upstream of the SNP site. After PCR amplification, the PCR products undergo the sequencing reaction; this only contains labeled ddNTPs for elongation which will get incorporated into the SNP site for detection.[41]

ii) Second Generation Sequencing

There are several different second-generation methods that have been developed; each of which takes advantage of a different aspect of the incorporation of dNTPs by *Taq* polymerase during elongation. One method is called Roche 454 sequencing (also known as pyrosequencing as pyrophosphate was used for detection). This method works by only applying a single dNTP type at a time during elongation; if the dNTP is complementary to the template, it will be incorporated in the growing strand and release pyrophosphate as a reaction by-product for detection. This detection is achieved by combining pyrophosphate with adenosine 5′-phosphosulfate (catalyzed by ATP sulfurylase), forming adenosine triphosphate which is a substrate for luciferase, and resulting in chemiluminescence. The amount of chemiluminescence detection is proportional to the number of bases incorporated in the case of repeating bases,[47] however, long repeats tend to be problematic for this method.[48]

Ion-torrent sequencing uses the same principle of only introducing a single dNTP at a time. However, instead of measuring incorporation of dNTPs through the pyrophosphate, the pH is measured instead since the reaction of dNTP incorporation acidifies the solution slightly. A semiconductor detector, instead of optics, is used to measure the pH.[49] This method suffers from the same limitation as pyrosequencing when it comes to long repeats.[50]

Illumina sequencing uses dNTPs labeled with a fluorophore at the 3’ OH end of the sequence, with each dNTP having a different fluorophore. These labeled dNTPs are introduced simultaneously and there will only be a single addition regardless of the sequence because the 3’ OH is blocked. The incorporated base is then recorded by a scanner and is followed by the fluorophore being cleaved off which exposes the 3’ OH to allow the next dNTP to be incorporated.[51] The benefit of this method compared to the pyrosequencing and ion-torrent methods is that even if there is a long repeat in the sequence, the Illumina method will always perform single base incorporation per cycle.[52] We adopted this sequencing method in the NGS work presented in Chapter 4.

iii) Third Generation Sequencing

All the sequencing methods that have been mentioned so far require PCR amplification of the DNA prior to the sequencing reaction. Third generation sequencing allows for unamplified sequences (including single DNA molecules) to be sequenced. One such

method is called Nanopore sequencing.[53] This method uses an electric field to pull the DNA molecule through a nanopore (which can either be a biological porin[54] or synthetic nanotube[55]). The sequence is determined by monitoring the electric field around the nanopore which fluctuates depending on each nucleotide of the DNA currently passing through the nanopore. This not only allows for the sequence to be identified but also for epigenetic analysis (e.g. DNA methylation) as methylated cytosines cause different fluctuations in the electric field than unmodified nucleotides.[56] This method is still under development since the fast rate of which the DNA passes through the nanopore results in a high error rate.[57]

d) Ginseng Species Differentiation

Ginseng has come to mean in colloquial terms the two herbal species *Panax ginseng* (Chinese ginseng) and *Panax quinquefolius* (American ginseng). American ginseng was discovered in the early 1700s by a Jesuit cleric. He heard of the herb in Asia and noticed that the Mohawks were using the herb, hence discovering the new species. This newly discovered ginseng became popular in China since it had its own unique set of medicinal effects, thus it was not a direct competitor of the Asian species.[58]

Ginseng is a small plant that grows in shaded areas (typically forest floors). The plant has five leaves and requires three years to reach maturity at which it will produce red berries. The most valuable part of the plant is its root as it contains the largest amount of ginsenosides, which are the medicinal compounds of the herb.[59] Ginseng products made from these roots have been used as herbal supplements in Traditional Chinese Medicine (TCM) for centuries. Globally, the ginseng industry produces (both *Panax* species) over 80 000 tons of ginseng products per year with an estimated value of $2 billion USD in 2013.[60]

Ginseng is a tetraploid plant with 48 chromosomes (12 chromosomes per haploid set).[61] The genomes size varies between *P. ginseng* and *P. quinquefolius* (3.12 Gbp and 4.8 Gbp, respectively).[62] Hybrids of the two ginseng species have been created (containing two copies of the *P. ginseng* genome and two copies of the *P. quinquefolius* genome). These hybrids are made because they have increased the ginsenoside content (specifically Rg2 and Rc) compared to the pure species. These hybrids have sterile pollens, however, there is no mention of the ovaries also being sterile.[63] Therefore, it is not possible for two hybrid plants to be crossed, however, it is theoretically possible for a hybrid ginseng (female) to be crossed with a pure species (male) although there is no report of such a cross-breeding being made.[64] This means that a ginseng sample will have one of three genotypes: pure *P. ginseng*, pure *P. quinquefolius* and 50:50 hybrid.

i) Dammarenediol-II Synthase

The enzyme involved in the production of the ginsenosides is called dammarenediol-II synthase (DS) which converts 2,3-oxidosqualene to dammarenediol-II (Figure 1.5).[65] Dammarenediol-II then undergoes hydroxylation by cytochrome P450s and glycosylation by glycotransferases at various positions[66] to form a variety of different ginsenosides.[65] Some examples of the ginsenosides and their aglycones (compounds remaining after the

glycosyl groups are replaced by H atoms) that are formed from dammarenediol-II are shown in Figure 1.6.

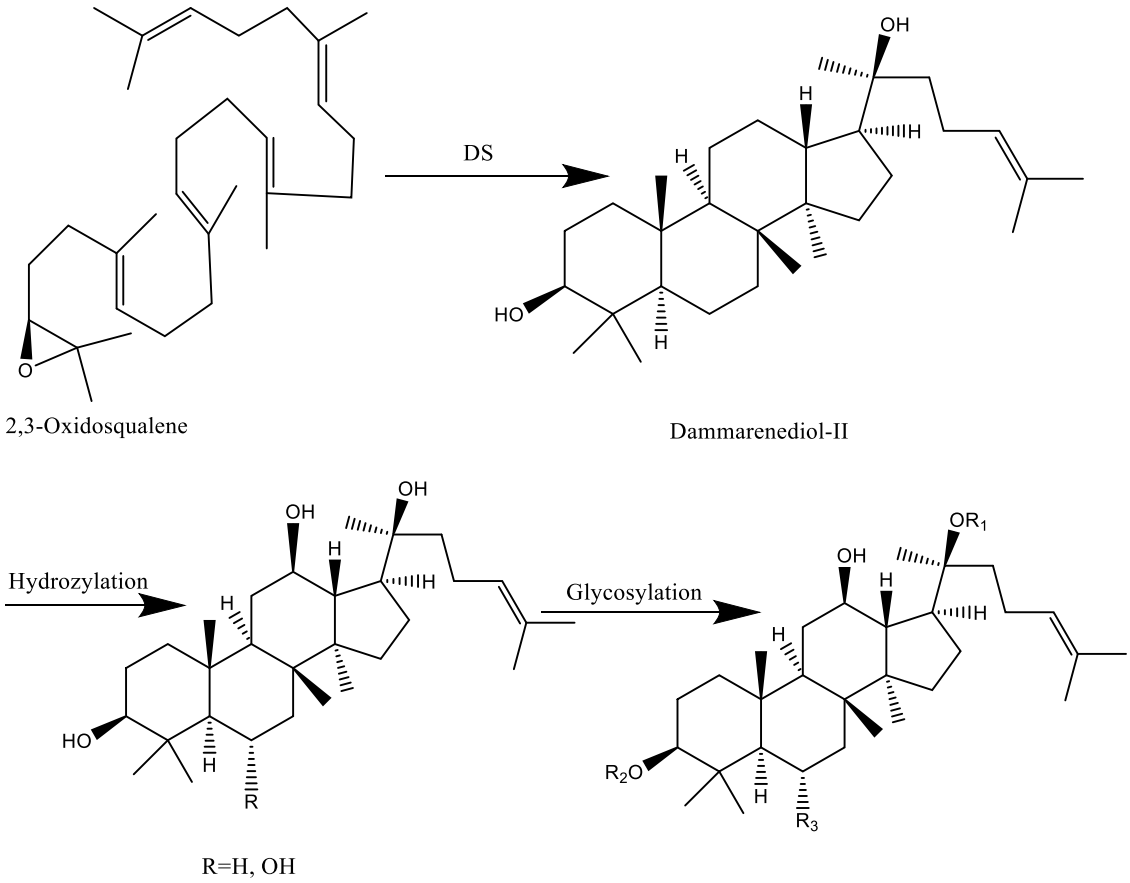
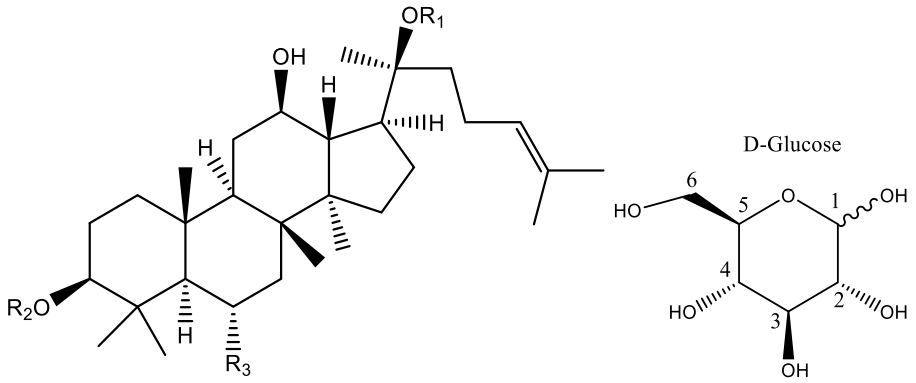


Figure 1.5: The conversion of 2,3-oxidosqualene to dammarenediol-II which is subsequently hydroxylated by ginseng cytochrome P450s and then glycosylated by glycotransferases.



Ginsenoside	R_1	R_2	R_3
PPD	H	H	H
Rb1	Glc^6-Glc	Glc^2-Glc	H
Rb2	Glc^6-Ara(p)	Glc^2-Glc	H
Rb3	Glc^6-Xyl	Glc^2-Glc	H
Rc	Glc^6-Ara(f)	Glc^2-Glc	H
Rd	Glc	Glc^2-Glc	H
Rg1	H	Glc^2-Glc	H
PPT	H	H	OH
Rh1	H	H	O-Glc
Re	Glc	H	O-Glc^2-Rha
Rg2	H	H	O-Glc^2-Rha
Rg3	Glc	H	O-Glc

Figure 1.6: The common ginsenosides found in *P. ginseng* and *P. quinquefolius* where Glc=D-glucose, Ara(p)=L-arabinose (pyranose form), Ara(f)=L-arabinose (furanose form) Xyl=xylose, Rha=L-rhamnose, PPD=protopanaxadiol and PPT=protopanaxatriol. Superscripts indicate the position (refer to image of D-glucose) of the glycosidic linkage.[67–72]

ii) Medicinal Effects of Ginseng

P. ginseng and *P. quinquefolius* have different ratios of these different ginsenosides. For example, *P. ginseng* has more Rg1, Rg2, Rh1, Rb2 and Rc while *P. quinquefolius* has more Rb1, Re and Rd.[73] In fact, *P. ginseng*/ *P. quinquefolius* may be chemically differentiated by the Rg_1/Re ratio.[74] This difference in the ginsenoside contents consequently results in the two species having different medicinal effects. In general, *P. ginseng* causes an increase in cardiovascular and neurological activity resulting in the person feeling energetic (the "hot" effect), while *P. quinquefolius* causes a decrease in cardiovascular and neurological activity which result in a calming feeling in a person (the "cool" effect).[73] Experimentally, the two ginsenosides Rg1 and Rb1 have been shown to be responsible for the neurological effect which subsequently influences the cardiovascular activity.[75]

P. ginseng has been also shown to have anticancer activity through the activity of the two ginsenosides Rg3 and Rh1. Since both ginsenosides are not present in significant quantities in *P. quinquefolius*, it does not have a significant anticancer effect.[76,77] Another pharmacological effect that varies between *P. ginseng* and *P. quinquefolius* is in their

interactions with other drugs. For example, *P. quinquefolius* has the ability to inhibit the blood thinner, warfarin[78] while *P. ginseng* will not inhibit it.[79,80] So, it is important for warfarin-taking patients to avoid *P. quinquefolius*. Underlying mechanism of the warfarin inhibition involves the ginsenosides enhancing the expression of hepatic cytochrome P450s which in turn results in an increase rate of metabolism and clearance of the warfarin. It is not clear which ones of the ginsenosides are responsible for this inhibition effect.[81]

During metabolism, the ginsenosides undergo deglycosylation to form the aglycones which reveal the hydroxyl groups at which the glycosyl moieties were attached and if the backbone is fully deglycosylated, protopanaxadiol (PPD) or protopanaxatriol (PPT) are formed.[82] These two metabolites have been shown to have medicinal effects themselves. Both PPD and PPT have been shown to inhibit proliferation and migration of cancer cells although the compounds are effective against many different cell lines.[83,84]

While the medicinal effects of these two species of ginseng are very different, their physical appearance is very similar especially considering that they are usually sold just as the root instead of the whole plant. This makes it difficult to authenticate the sample as either *P. ginseng* or *P. quinquefolius*. Furthermore, it is quite common for the root to be sold ground up into a powder instead of an intact root which is impossible to classify as *P. ginseng*, *P. quinquefolius* or a hybrid of the two species[63] by physical appearance. However, chemical, and genetic authentication will always work regardless of what physical form the ginseng sample is in.

iii) Methods of Chemical Authentication of Ginseng

Chemical authentication of ginseng usually consists of analyzing the ginsenoside content of ginseng by LCMS.[85,86] A common metric that is used for determining the species is the Rg1/Re ratio, with a high ratio indicating that the sample is *P. ginseng* and a low ratio indicating that it is *P. quinquefolius*.[74] Capillary electrophoresis, with UV detection, has also been used for ginsenoside characterization.[87] Another category of biomolecules in ginseng that can be used for species identification are polysaccharides.[88] *P. ginseng* contains a greater proportion of acidic polysaccharides (containing uronic acid) while *P. quinquefolius* contains more neutral polysaccharides.[89]

The major downside of chemical authentication is that the ginsenoside composition can vary not only as a result of the species but also of the growing environment, the age of the ginseng,[90] possible adulteration[91] and the specific part of the root that is used for analysis.[92]

iv) Methods of Genetic Authentication of Ginseng

One way to perform genetic authentication is to perform DNA sequencing. It has the advantage of being able to detect any type of genetic variations (point mutation, insertion, deletion, etc.). Both the nuclear genomes[93] and chloroplast genomes[94] of ginseng have been sequenced and variations have been found in different species. The sequencing

method does, however, come at the cost of requiring highly pure genetic material and costly sequencers.

A common method that has been used for genetic authentication of ginseng is by using the intron targeting SNP (ITS) markers. This method works by making different length of PCR products depending on whether the sample is *P. ginseng* or *P. quinquefolius*. This is performed by designing one forward primer with the SNP base at the 3' end of the primer so that the PCR product is only produced if the template is *P. ginseng* and a second forward primer is designed for a second SNP site using the same criteria except that a PCR product is produced only if the sample is *P. quinquefolius*. The reverse primer is a universal primer for both species.[95] Various genes (e.g. auxin repressor gene,[95] mitochondrial cytochrome C oxidase subunit 2 gene,[96] dammarenediol-II synthase gene[97–99]) have also been studied using the ITS markers. The PCR products must be analyzed afterwards by gel electrophoresis in order to determine their lengths. Work has been done in order the perform capillary electrophoresis in a microfluidic chip, but this method still requires an external device in order to control the electric potential of the process.[100]

Another method for ginseng species authentication is to perform PCR with random sequence primers which in turn generate random length PCR products which vary between *P. ginseng* and *P. quinquefolius*.[101,102] A variant of this method has been developed where genomic samples are digested into fragments with a specific restriction enzyme (e.g. EcoRI) followed by the ligation of adaptors for universal primer binding. The fragments are subsequently amplified by PCR and analyzed on a gel.[103] These methods, however, require not only gel electrophoresis but also authentic samples that are used as reference standards.

One way to do the authentication of ginseng species without requiring a system for gel electrophoresis is to use real time PCR (rt-PCR). The PCR primers are designed to hybridize to conserved regions of a locus and the quantification probes hybridize to a SNP site within the amplified region. Since the probe hybridization is done at an elevated temperature (60°C), the duplex will only form if there is a match at the SNP site.[104] This technique will be used in this thesis except that instead of analyzing the PCR product through rt-PCR, the PCR strand-probe duplexes are immobilized on a solid surface and quantified using fluorescence. When the PCR strands (or targets) are being analyzed, they will be classified into one of three possible genotypes: *P. ginseng*, *P. quinquefolius* or a hybrid of the two species.

v) Research Objectives

The objective of this thesis is to authenticate various *Panax* species. This will be achieved by analyzing different methods of forming probe-target complexes. Here, the probe is a short oligonucleotide (20 nt to 34 nt), while the target is the PCR strand or long oligonucleotides (from 126 nt to 694 nt). Specifically, three different methods will be examined: 1) stop flow probe-target method, 2) target-probe method, and 3) three-strand complex method.

Method 1) is the conventional method that was reported previously[42] and is thus used here to verify that the probe and target designed will hybridize well with each other.

The conventional Method 1 is to immobilize the short probe and to hybridize to it the long PCR strand. It is hypothesized that it would be easier for the short length probe strand to hybridize to the immobilized long length PCR strand than the conventional method. Thus, two different methods of immobilizing the PCR strand will be examined. One method involves immobilization of the PCR strand directly to the surface (Method 2) and the other involves indirect immobilization of the PCR strand through hybridization to a directly immobilized capture strand (Method 3). For these methods, after pumping the reagent solutions into the microchannels, hybridizations are conducted under the SF conditions.

Other than SF hybridizations, there has also been interest in developing the dynamic hybridization where a centrifugal force is used to pump the solution through the channels while using the microfluidic chip with straight channels. Dynamic hybridizations have been tested on the three methods mentioned above.

The development of various methodologies will be discussed. Different lengths of probes and PCR products as well as different SNP sites within the PCR products will be used when examining these different techniques.

The above methods are nucleic acid tests (NATs) based on hybridization of the target strand that is externally amplified using PCR. I would like to develop a nucleic acid amplification test (NAAT). Instead of combining method 3 with the isothermal amplification method called helicase-dependent amplification (HDA) that was previously developed in Dr. Li's group,[105] I developed a brand new method as follows.

Developing an on-chip isothermal amplification using the microfluidic chips was another goal of this thesis. The materials used to make the chip create major problems in terms of water evaporation and gas evolution when considering thermophilic (41-122 ºC) amplification methods (e.g. HDA[106]). An isothermal mesophilic (20-45 ºC) method has been developed by Kausar *et al* which uses T4 DNA ligase in a ligation chain reaction called lesion-induced DNA amplification (LIDA).[107,108] This reaction not only allows for the amplification of a target sequence but also detection of the amplified product without requiring probe hybridization. This method, originally developed in the solution phase, will be adapted to be used as a solid-phase method in the chip.

This solid-phase LIDA method is used to authenticate ginseng samples (powder and root). The results are compared with those obtained from next-generation sequencing (NGS). Details of this method are described in Chapter 4.

Chapter 2 Material and Methods

a) Materials

i) Reagents and Solvents

Photoresist SU-8 (2035) and SU-8 developer solution were purchased from MicroChem, Newton, MA. Silicon wafers were purchased from University Wafer, South Boston, MA, and the Sylgard^{TM} 184 Silicone Elasomer Kits (used to make the PDMS slabs) were purchased from Dow Silicon Corp., Midland, MI.

95% ethanol was purchased from Commercial Alcohols, isopropanol was purchased from Anachemia, sodium borohydride was purchased from Caledon Laboratories Ltd., and sodium cyanoborohydride, 3-(aminopropyl)triethoxysilane and 25% glutaraldehyde were purchased from Sigma Aldrich.

The PCR reagents (*Taq* polymerase, PCR buffer, $MgSO_4$ and dNTPs) was purchased as a kit from ABM and the T4 ligase was purchased from New England Biolabs (NEB) and Thermo Fisher (TF).

The genomic DNA was extracted using a DNeasy Plant Mini Kit (including buffer AP1, AW1, AW2, AE, QIAshredder spin column, DNeasy Mini spin column, RNase A). PCR products were purified using a QIAquick PCR Purification Kit (including buffer PB, PE, QIAquick spin column) from Qiagen.

ii) Buffers

Immobilization Buffer: 43.8 g of NaCl and 6.30 g of $NaHCO_3$ were dissolved in 500 mL of deionized (DI) water.

20× Phosphate Buffer Saline (PBS): 14.2 g of Na_2HPO_4, 2.45 g of KH_2PO_4, 80.0 g of NaCl and 2.00 g of KCl were dissolved in 500 mL of DI water.

1× PBS: 25 mL of 20× PBS were combined with 475 mL of DI water.

10× Saline Sodium Citrate (SSC): 43.8 g of NaCl and 22.1 g of sodium citrate dihydrate were dissolved in 400 mL of DI water. The solution was then acidified to pH 7.0 using 0.1 M HCl followed by the addition of more DI water to make a total volume of 500 mL.

1% Sodium Dodecyl Sulfate (SDS): 150 mg of SDS were dissolved in 15 mL of DI water.

1× PBS/Tween: 1.5 µL of Tween-20 was mixed with 1.0 mL of 1× PBS.

b) Chip Production and Glass Functionalization

i) Si/SU-8 Mold for Chip Fabrication

A clean silicon wafer was put in a glass crystallizing dish and the silicon surface was cleaned using piranha solution (35 mL conc. H_2SO_4, 15 mL 30% H_2O_2) heated to 95°C for 15 min. Take extreme care with the use of piranha solution because it is highly

corrosive. Afterwards, the silicon wafer was washed using DI water and the wafer was dried with air.

The wafer was then placed on the spin coater (WS-400, Laurell Technologies Corp. North Wales, PA) and the suction was turned on to hold the wafer in place. SU-8 was applied to the center of the wafer so as to cover approximately 25% of the wafer surface area. A pipette tip was used to move any air bubbles to the edge of the SU-8 puddle. The wafer was then spun first at 500 rpm for 10 s to spread out the SU-8 puddle followed by 4000 rpm for 30 s to form a thin SU-8 layer.

The SU-8 coated wafer was next placed in a clean glass petri dish and placed on a 65°C hot plate for 3 min, followed by placing the dish in a 95°C oven for 7 min. Once the hot baking was completed, the wafer was cooled to room temperature.

The photomask, which contained the chip design pattern, was then placed on top of the SU-8 layer, with the less shiny side of the mask down, and the assembly was placed in the UV lamp (Model LS-150-3, Bachur & Associates, San Jose, CA). A glass plate was placed on top to hold the photomask down. The SU-8 on the wafer was irradiated with 365 nm UV light for 5 s at 175 mA/cm^2. After the UV exposure, the glass plate and photomask were removed and the wafer was placed back in the glass petri dish, baked on the 65°C hot plate for 1 min followed by the 95°C oven for 7 min. After the baking, the wafer was cooled to room temperature.

Next, the wafer was submerged in SU-8 developer solution and the dish was swirled for a few minutes. This was followed by rinsing the wafer with fresh SU-8 developer, then isopropanol and then by drying the wafer with nitrogen gas. The preparation of the SU-8 mold on the wafer was completed. Silicone sealant was applied along the circumference of the mold which was allowed to cure overnight.

ii) PDMS Slab Casting

A PDMS slab containing 16 parallel microchannels (150 µm X 35 µm cross section) was made by pouring a mixture of PDMS elastomer and curing agent (10:1 v/v) into a Si/SU-8 mold and allowing the polymer to cure for 3 days (Figure 2.1). The first PDMS slab

made from the mold was sacrificed in order to measure the cross-sectional dimensions of the channels.

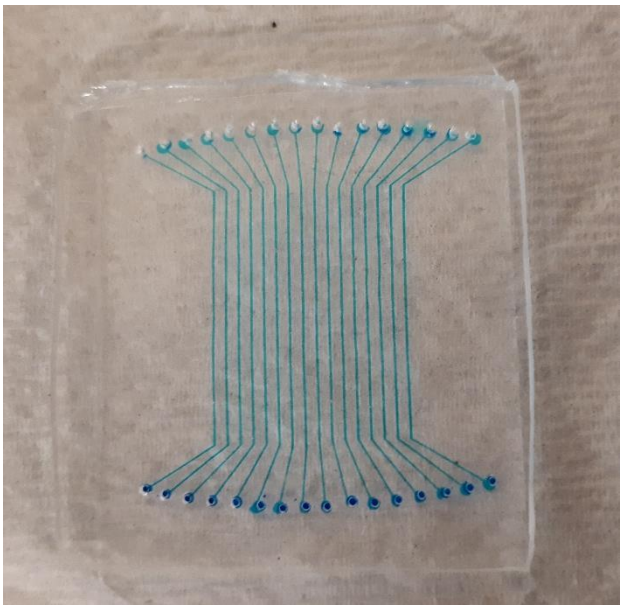


Figure 2.1: Image of the 16 channel PDMS slab on a glass slide. The channels have been filled with blue dyed solution for illustration.

iii) Aldehyde Glass Functionalization

Glass chips were made using Fisher 50 mm X 75 mm glass microscope slides. The slides were washed with Sparkleen solution, LiquiNOX solution, DI water, 95% ethanol and DI water and then dried using air. These slides were further cleaned with piranha solution (70 mL concentrated H_2SO_4, 30 mL 30% H_2O_2) for 15 min at 95°C. The slides were subsequently rinsed with DI water and dried using air. Next, the slides were amine-functionalized by immersing them in a (3-aminopropyl)triethoxysilane (APTES, see Figure 2.2a) solution (2 mL APTES, 98 mL 95% ethanol) for 20 min at room temperature (rt). The slides were subsequently washed with 95% ethanol and dried using nitrogen gas. Next, the slides were placed in an oven for 1 h set to 120°C. After cooling to rt, the slides were aldehyde-functionalized by immersing them in a glutaraldehyde solution (5 mL 20× PBS, 75 mL DI water, 20 mL 25% glutaraldehyde) for 1 h (Figure 2.2b) at 4°C. Afterwards, the slides were washed with DI water and dried using nitrogen gas. The slides were stored at 4°C for later use.

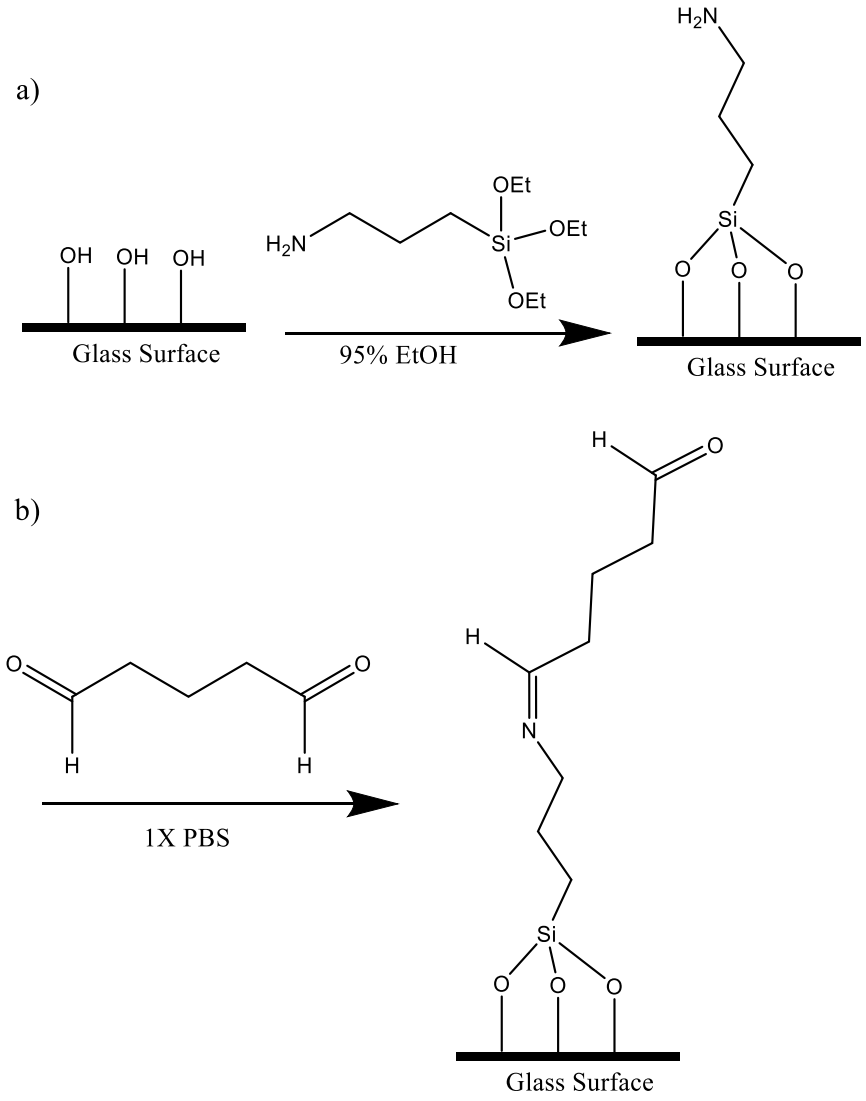


Figure 2.2: The functionalization of the glass surface with aldehyde groups. a) The hydroxyl surface of the glass is functionalized with amine groups using (3-aminopropyl)triethoxysilane. b) The amine surface is converted to an aldehyde surface using glutaraldehyde.

c) Ginseng Probe and Primer Design

i) Sequence Information

The sequencing of the DS gene has been previously reported (accession number JN596111).[97,98] The section of the DS sequence that will be used in this thesis is shown in Figure 2.3. This is an AT rich sequence with a GC content of only 32%. This means that the probes and primers will have to be long (*ca.* 30 nt), i.e. longer than the typical length between 20 to 25 nt to maintain the Tm values. Shorter length versions of the probes (20 nt) have also been designed despite the low GC content to test whether the probe length itself has an effect.

ii) PCR Primer Sets

Five different SNP sites have been identified as shown in Figure 2.3. Three different PCR product lengths were made using different combinations of (forward and reverse)

primers (Table 4): P3, P6 (694 bp product, PCRa), P8, P7 (226 bp product, PCRb) and P8', P7 (126 bp product, PCRc). PCRa and PCRb encompass all five SNP sites while PCRc only encompasses the three single SNP sites (N1, N2, N3, see Figure 2.3). The primers are designed using the software DINAMelt[29] based on design rules such as oligonucleotide length, GC content, melting temperature, absence of secondary structure. All DNA oligonucleotides used were purchased from Integrated DNA Technologies (IDT). PCRa will function as a positive control as this PCR product has been previously reported.[97,99] The two shorter length PCR product (PCRb and PCRc) are made to examine the effects of intramolecular folding on the extent and selectivity of the strand hybridizing to the probe. The longer one of the two PCR products (i.e. PCRb, 226 bp) has more intramolecular folding of the strand that hybridizes to the probe, compared to the shorter (PCRc, 126 bp) product.

Table 4: List of all primers and their functionalization. Note that the P6 and P7 primers each have two different functional groups on the same primer sequence (amine for target-probe method; biotin for probe-target method).

Oligonucleotide Name	Functional Group	Sequence (5' to 3')	Purpose
A-P7 Primer	5'-Amine	ATGTTTGT TACTCCCTCC GTT	reverse primer for PCRb and PCRc
B-P7 Primer	5'-Biotin	ATGTTTGT TACTCCCTCC GTT	reverse primer for PCRb and PCRc, biotin-labeled
A-P8 Primer	5'-Amine	TACTCCAT TCGTTCCATA ATATTTGAT	PCRb forward primer
A-P8' Primer	5'-Amine	TACAGTGAT AATTAAATAT TGTAACTATC TAA	PCRc forward primer and PCRa,b and c (antisense) capture strand
A-P3 Primer	5'-Amine	ATGCAGTTACAACCGCTGTG	PCRa forward primer
B-P6 Primer	5'-Biotin	TATAGGATCCCCATCCACCA	PCRa reverse primer, biotin-labeled
P6 Primer		TATAGGATCCCCATCCACCA	PCRa reverse primer, unlabeled

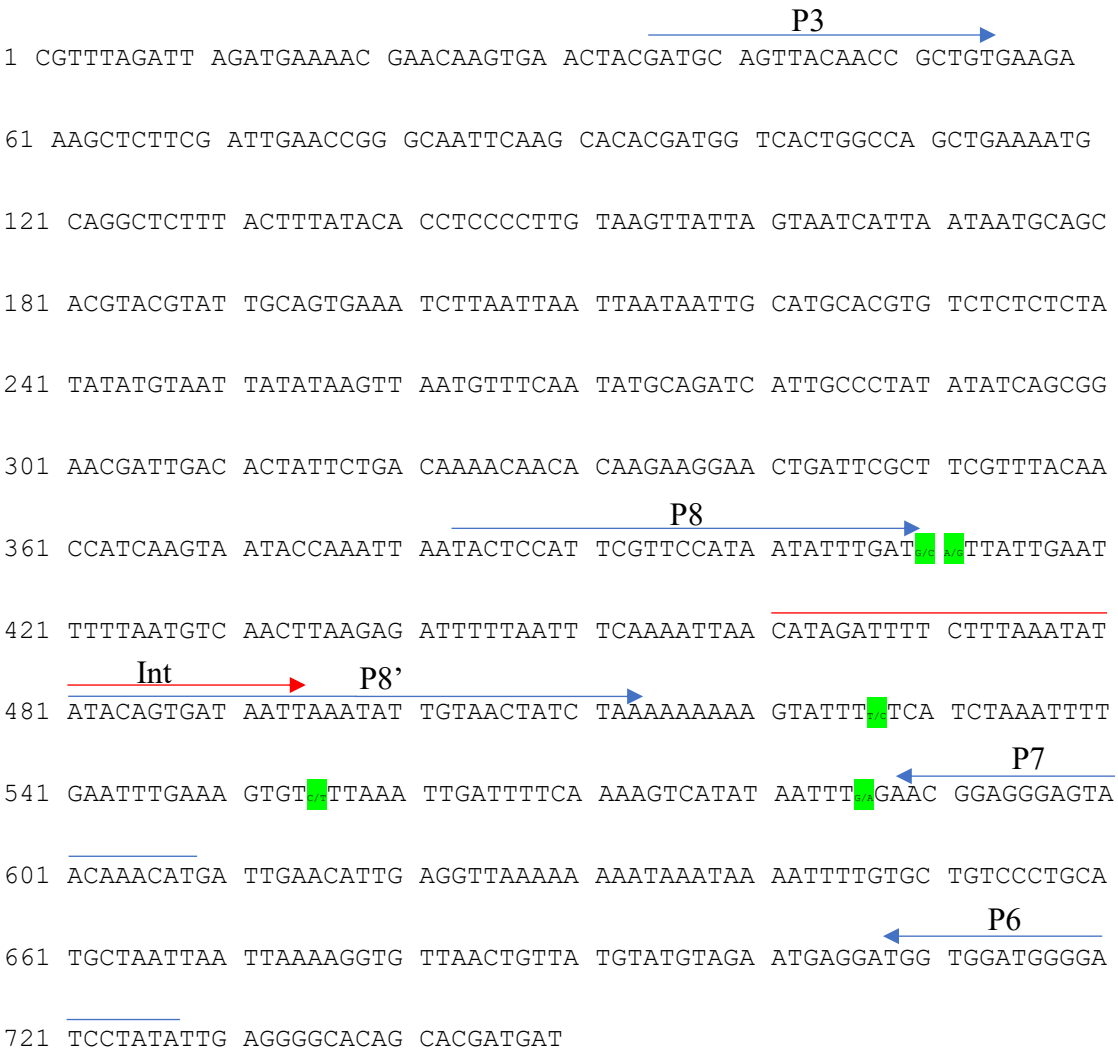
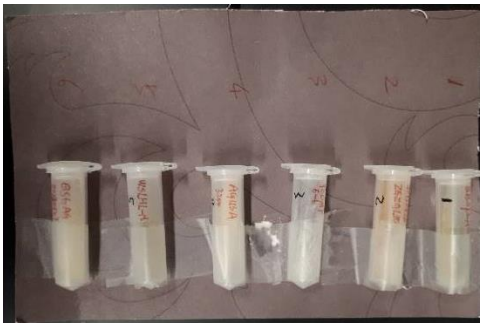


Figure 2.3: A section of the DS gene sequence (sense strand, 5' to 3'). The nucleotides in green represents the five SNP sites; the first character being the one found in *P. ginseng* and the second being that in *P. quinquefolius*. The three single SNP sites have been sequentially labeled as N1, N2 and N3, at positions 527, 555 and 586, respectively. The tail of the arrows represents the 5' end and the head of the arrow represents the 3'end of the primers (blue) and the Int probe (red). The reverse primer sequences are complementary to the sequence presented.

a)

c)



b)

d)

Figure 2.4: a) A whole ginseng root (AuChG), b) the central section of a root sample (AmG2) (left) and cross-sectional slice of the root (AmG) (right), c) six ginseng root powder samples (X1-6) and d) AmG3 root sample.

iii) Genomic DNA Extraction

Six ginseng samples (Figure 2.4c) were received (as a fine powder) from Macan Biotechnology Ltd. (Macau). A certified root was received (Figure 9a), and a small dendrite was cut and crushed into powder. Additional root samples were also received (Figure 2.4b, d) which were crushed into a fine powder using a mortar & pestle. The genomic DNA was extracted from these samples using a Qiagen plant DNA extraction kit based on the procedure (Quick-Start Protocol, DNeasy® Plant Mini Kit, March 2016) provided by the kit. In short, *ca.* 90 mg of each sample (as a fine powder) was placed in a 600 µL centrifugal tube and suspended in 400 µL of Buffer AP1 and 4 µL of RNase A. This tube was vortexed and incubated at 65°C for 10 min. The tube was inverted three times during the incubation. Next, 130 µL Buffer P3 was added, mixed by pipetting, and incubated on ice for 5 min. The mixture was then centrifuged for 5 min at 13,000 rpm. The lysate was pipetted into a QIAshredder spin column and centrifuged for 2 min at 13,000 rpm. The flow-through was transferred to a new 1.5 mL centrifugal tube and 750 µL of Buffer AW1 was added and mixed by pipetting. 650 µL of the mixture was transferred to a DNeasy Mini spin column and spun for 1 min at 8000 rpm. The flow-through was discarded and the process was repeated for the remainder of the sample. The spin column was then placed in a new 2 mL collection tube and 500 µL of Buffer AW2

was added and then centrifuged for 1 min at 8000 rpm. The flow-through was discarded and another 500 µL of Buffer AW2 was added and centrifuged for 2 min at 13,000 rpm. The spin column was transferred to a new 1.5 mL centrifugal tube and 100 µL Buffer AE was added and centrifuged for 1 min at 8000 rpm. Another 100 µL Buffer AE was added to the spin column and centrifuged for 1 min at 8000 rpm.

iv) PCR Amplification

A PCR reagent kit (ABM) was used for the PCR. The PCR mixture consisted of: 10× PCR buffer (with 15 mM $MgCl_2$) (5 µL), forward primer (1.5 µL @ 10 µM), reverse primer (1.5 µL @ 10 µM), genomic DNA (10 µL) or 0.5 µL (30 ng) purified PCR product, $MgSO_4$ (1.0 µL @ 25 mM), dNTPs (1.0 µL @ 10 mM), *Taq* polymerase (1.0 µL @ 5 U/µL) and deionized (DI) water (29 µL for genomic templates or 38.5 µL for PCR products templates) (final volume 50 µL).

The thermocycling was done in a Techne ^{3}Prime thermocycler. The program for PCR mixtures consisted of a 3 min 94°C initial denaturation, followed by 30 cycles of 30 s 95°C denaturation, 30 s primer annealing (52°C for PCRa, 50°C for PCRb and PCRc) and 30 s 72°C elongation, and was subsequently followed by a final elongation at 72°C for 3 min.

The PCR mixtures were subsequently purified using a Qiagen PCR purification kit (Quick-Start Protocol, QIAquick® PCR Purification Kit, July 2018) In short, 250 µL of Buffer PB was added to the 50 µL of PCR mixture and mixed by vortexing. The mixture was then transferred to a QIAquick spin column and centrifuged for 1 min at 13,000 rpm. The flow-through was discarded and 750 µL of Buffer PE was added to the spin column and centrifuged for 1 min at 13,000 rpm. The flow-through was discarded and the spin column was centrifuged again at 13,000 rpm for 2 min. The spin column was then transferred to a clean 1.5 mL centrifugal tube and the PCR product was eluded using 30 µL of DI water centrifuged for 1 min at 13,000 rpm.

The PCR product was quantified using a Thermo Scientific Nanodrop 1000. The DNA was quantified based on the absorbance at 260nm, with its ratio to the 280 nm absorbance noted.

In addition, asymmetric PCR products (one strand is in excess) were made consisting of: 10× PCR buffer (with 15 mM $MgCl_2$) (5 µL), P8 forward primer (0.67 µL @ 10 µM), P7 reverse primer (2.5 µL @ 10 µM), purified PCR product DNA (0.5 µL, 30 ng), $MgSO_4$ (1.0 µL @ 25 mM), dNTPs (1.0 µL @ 10 mM), *Taq* polymerase (1.0 µL @ 5 U/µL) and deionized (DI) water (38.33 µL) (final volume 50 µL). The asymmetric PCR was done using the same thermocycler conditions as before, as were similarly used for the corresponding primer sets in the genomic PCR products as well as the subsequent purification and quantification.

v) Probe and Target Strands for Detection of Three SNP Sites

The three SNP sites (N1, N2, N3) seen in Figure 2.3 were chosen for comparing the efficiency of hybridization of duplexes with perfect matches verses that of duplexes with

a 1-bp mismatch. The N1 probe strands shown in Table 5 hybridize close to the 3' end of the PCR antisense strand (with the sense sequence shown in Figure 2.3); the N2 probe strands hybridize to the central region of the PCR strand and the N3 probe strands hybridize near the 5' region of the PCR strand.

This means that when the probe-target method (conventional Method 1) is being used, the PCR product strand (target) hybridizes with its 3' end facing down towards the solid surface. Thus, if an N1 probe is used, most part of the PCR strand will face upwards into the solution and the strand will have minimal steric effect with the solid surface (Figure 2.5a, left). On the other hand, the N3 probe will position a large portion of the 3' end of the PCR strand downwards towards the solid surface resulting in a lot of steric hindrance (Figure 2.5a, right).

When the target-probe method (Method 2) is used, the PCR strand is immobilized directly onto the surface (at its 5' end). In this case, the N1 position will be the farthest from the surface thus the most accessible to hybridization while the N3 position will be the closest to the surface and the least accessible (Figure 2.5b).

Finally, in the three-strand method (Method 3), when the PCR strand is immobilized using a capture strand, the three SNP sites will be in reverse order to Method 2 (i.e. N3 most exposed, N1 least exposed, Figure 2.5c).

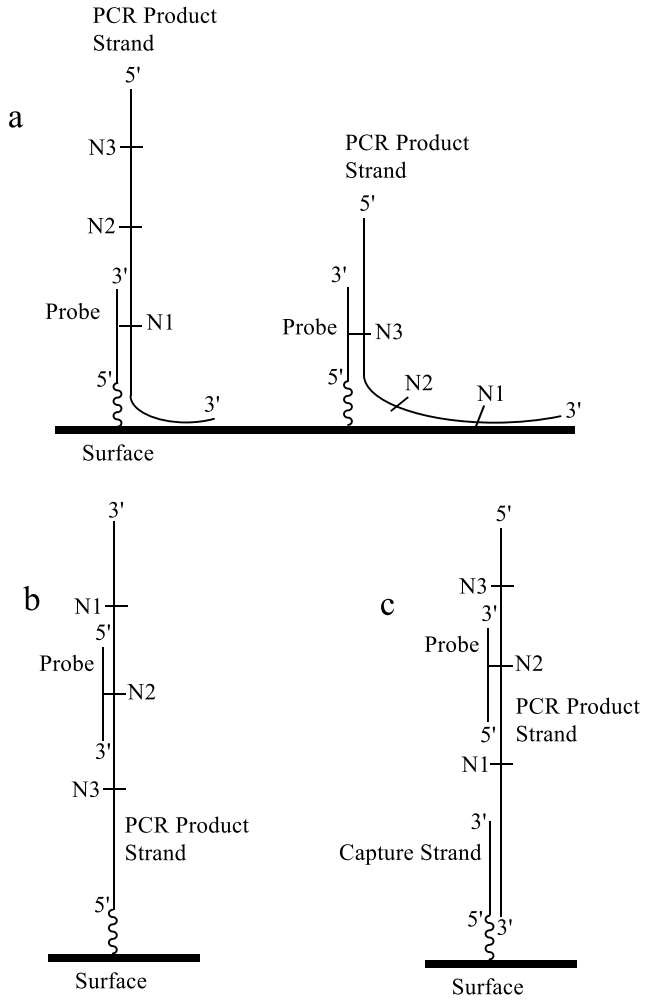


Figure 2.5: The orientation of the SNP sites on the PCR strands (antisense) using different techniques a) conventional probe-target Method 1 using an N1 probe (left) and N3 probe (right) b) target-probe Method 2 c) capture-strand-target-probe Method 3.

Table 5: List of all oligonucleotide strands with their respective functionalization, sequence, and purpose in the hybridization experiments. Probes with names containing N1, N2 and N3 hybridize to the SNP sites (N1, N2 and N3, respectively) shown in Figure 2.3. The Probe-Target method is the conventional Method 1 used that involves immobilizing a short probe and hybridizing it with a longer length PCR strand (Figure 2.5a) while the Target-Probe method is the new Method 2 being developed for immobilizing the long PCR product and hybridizing it with a shorter-length probe (Figure 2.5b). Method 3 is the three-strand method that used a capture strand (immobilized) to capture the PCR strand, followed by hybridizing it with the detection probe (Figure 2.5c).

Strand Name	**Functionalization**	**Sequence (5' to 3')**	**Purpose**	**Method**
B-N1G	5'-Biotin (or B)	C TAAAAAAAAA GTATTTTTCA TCTAAATTTT GAA	Detection probe	2, 3
B-N1Q	5'-Biotin	C TAAAAAAAAA GTATTTCTCA TCTAAATTTT GAA	Detection probe	2, 3
B-N2G	5'-Biotin	GAATTTGAAA GTGTCTTAAA TTGATTTTCA A	Detection probe	2, 3
B-N2Q	5'-Biotin	GAATTTGAAA GTGTTTTAAA TTGATTTTCA A	Detection probe	2, 3
A-Gin	5'-Amine (or A)	TGAAAATCAA TTTAAGACAC TTTCAAATTC AAAATTTAGA TGAAAAATAC TTTTTTTTTA	Target for N1/N2 detection probes	2
A-Quin	5'-Amine	TGAAAATCAA TTTAAAACAC TTTCAAATTC AAAATTTAGA TGAGAAATAC TTTTTTTTTA	Target for N1/N2 detection probes	2
A-N1G	5'-Amine	C TAAAAAAAAA GTATTTTTCA TCTAAATTTT GAA	probe	1
A-N1Q	5'-Amine	C TAAAAAAAAA GTATTTCTCA TCTAAATTTT GAA	probe	1
A-N2G	5'-Amine	GAATTTGAAA GTGTCTTAAA TTGATTTTCA A	probe	1
A-N2Q	5'-Amine	GAATTTGAAA GTGTTTTAAA TTGATTTTCA A	probe	1
B-Gin	5'-Biotin	TGAAAATCAA TTTAAGACAC	Detection target for	1

		TTTCAAATTC AAAATTTAGA TGAAAAATAC TTTTTTTTTA	N1/N2 probes	
B-Quin	5’-Biotin	TGAAAATCAA TTTAAAACAC TTTCAAATTC AAAATTTAGA TGAGAAATAC TTTTTTTTTA	Detection target for N1/N2 probes	1
A-Int	5’-Amine	CATAGATTTT CTTTAAATAT ATACAGTGAT AATT	Non-specific probe, Internal standard	1
A-tN2G	5’-Amine	TGAAAGTGTCTTAAATTG AT	20 nt probe	1
B-tN2G	5’-Biotin	TGAAAGTGTCTTAAATTG AT	Detection probe (20 nt)	2
A-tN1Q	5’-Amine	AAAAGTATTTCTCATCTA AA	20 nt probe	1
B-tN1Q	5’-Biotin	AAAAGTATTTCTCATCTA AA	Detection probe (20 nt)	2
A-N3Q	5’-Amine	CATATAATTTAGAACGGA G	20 nt probe	1
B-N3Q	5’-Biotin	CATATAATTTAGAACGGA G	Detection probe (20 nt)	2
A-LN3Q	5’-Amine	AAAGTCATATAATTT AGAACGGAGGGAGTA	30 nt probe	1
3’A-P8’	3’-Amine	TACAGTGATAATTAAATA TTGTAACTATCTAA	3’ Capture strand for PCRb	3

In addition to the probes (indicated by N1, N2 or N3 in their names), Table 5 also contains the two oligonucleotides Gin and Quin that are complementary to both N1 and N2 probes. The use of these two oligonucleotides is to function as positive controls in hybridization experiments.

d. Sample Preparation for DNA Sequencing

All sequencing was performed by Genewiz. Two different types of sequencing methods were used: i) Sanger sequencing and ii) Illumina sequencing.

i) Sanger Sequencing

10 µL of 1 ng/µL PCR product (2 ng/µL for 694 nt products) solutions (based on Nanodrop readings) were prepared in a centrifugal tube along with 5 µL of 5 µM P8 primer (P3 primer for 694 nt products) in a separate centrifugal tube as per instructions on the Genewiz website for Sanger sequencing.

ii) Illumina Sequencing

PCR products (226 nt) derived from 14 different ginseng samples were used as templates for preparing the sequencing library. Each of these samples was amplified by two rounds of PCR using the PCR procedure described in section c)iv). First round of PCR was conducted using the PCR1 primer pair (Table 6). The products were then purified and used as templates for a second round of PCR using the PCR2 primers. Here, each ginseng sample receive the adapter sequence and a different combination of indices in the PCR2 forward and reverse primers. The adapter sequence allows for bridge amplification in the Illumina sequencing protocol.[109]

The red region in the PCR2 forward primer is complementary to the 5' end of the PCR1 forward primer (the left half of the toehold). The blue region in PCR2 reverse primer is complementary to the 5' end of the PCR1 reverse primer (the left half of the toehold). The purpose of using the PCR1 primers (specific for locus) and PCR2 primers (generic, for adapter and indices) as opposed to a single set of primers is two-fold: 1) It is hard to synthesize the full length of the adapter sequence with the index because of oligonucleotide length limitations. 2) This allows the PCR2 primers to be used as generic primers (they can be used for any species/locus). Only the PCR1 primers have to be custom made for the specific locus of interest.

Table 6: List of all PCR primers used for the Illumina sequencing. In the PCR1 primers, the bold sections are locus-specific that are the same as the P8' and P7 primers; the italic sections form toeholds on the PCR1 product during the first-round amplification. The complementary sections of these parts of the toeholds (red or blue) hybridize to the 3' end of the PCR2 primers (red or blue) in the second-round amplification process which forms a set of adapters. These adapters contain the universal primers (italic) and indices (underlined) in the PCR2 primers which are used in bridge amplification in Illumina sequencing.

PCR1 Primers	
Forward PCR1 Primer	*TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG***TAC AGTGATAATTAAATATTGTAACTATCTAA**
Reverse PCR1 Primer	*GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG***AT GTTTGTTACTCCCTCCGTT**
PCR2 Forward Primers	
Forward PCR2 Primer 1	*AATGATACGGCGACCACCGAGATCTACAC*AGCGCT TCGTCGGCAGCGTC
Forward PCR2 Primer 2	*AATGATACGGCGACCACCGAGATCTACAC*GATATC TCGTCGGCAGCGTC
Forward PCR2 Primer 3	*AATGATACGGCGACCACCGAGATCTACAC*CGCAGA TCGTCGGCAGCGTC
Forward PCR2 Primer 4	*AATGATACGGCGACCACCGAGATCTACAC*TATGAG TCGTCGGCAGCGTC
PCR2 Reverse Primers	
Reverse PCR2 Primer 1	*CAAGCAGAAGACGGCATACGAGAT*GTGAATGTC TCGTGGGCTCGG
Reverse PCR2 Primer 2	*CAAGCAGAAGACGGCATACGAGAT*ACAGGCGTC TCGTGGGCTCGG
Reverse PCR2 Primer 3	*CAAGCAGAAGACGGCATACGAGAT*CATAGAGTC TCGTGGGCTCGG
Reverse PCR2 Primer 4	*CAAGCAGAAGACGGCATACGAGAT*TGCGAGGTC TCGTGGGCTCGG

After purification of the fourteen PCR2 products, these products were mixed in equimolar amounts (according to Nanodrop readings) to have a cumulative DNA concentration of 4 ng/µL in DI water. 10 µL of the mixture was sent to Genewiz for Illumina sequencing (MiSeq).

e. Thermodynamics Softwares (DINAMelt, Tm, Mfold)

All DNA hybridization thermodynamics (DINAMelt) and intramolecular folding thermodynamics (Mfold)[29] were conducted at 22°C with 150 mM Na^+ for all probes and PCR products. All primers hybridizations were conducted at 50 mM Na^+ and 3 mM Mg^{2+}. All other parameters were kept at their default settings. The thermodynamics of hybridization of the LIDA fragments were conducted at 37°C with 10 mM Na^+ and 10 mM Mg^{2+}.

Tm values are calculated by DINAMelt and measured using UV absorbance spectroscopy (see Appendix D.

f. Microfluidic Chip Experiment Procedure

Hybridization experiments were conducted on the PDMS-glass chips based on the NAT method shown in Figure 1.4. There are two steps involved, namely i) immobilization and ii) hybridization, which are described in details as follows.

i) Strand Immobilization

In the first step of NAT experiments, we performed strand immobilization. We either immobilized single-stranded oligonucleotides (short probes or long targets) or asymmetric PCR strands.

a) Immobilization of Oligonucleotide Strands

1.0 μL of the stock oligonucleotide solution (500 μM) was mixed with 19.0 μL of immobilization buffer forming a 25 μM solution. The 16 channel PDMS slab was washed with: sparkleen solution, LiquiNOX, DI water, 95% ethanol, and DI water, dried with air and sealed to the aldehyde glass slide such that the channels run horizontally. Then 1.0 μL of the 25 μM oligonucleotde solutions were applied to the channel wells and suction was used to pull the solution into the channels. Formation of Schiff bases occurred between the amine groups on the oligonucleotides and the aldehyde groups on the glass surface. The solution was incubated at rt for 1 h and then removed using suction. The channels were then washed with immobilization buffer followed by the removal of the PDMS slab from the glass slide. To allow for chemical reduction of the Schiff bases, the glass slide was submerged in $NaBH_4$ solution (~2.5 mg/mL $NaBH_4$, 0.75× PBS, 23.75% ethanol) for 15 min, washed with 1× PBS and dried with nitrogen gas. The successful probe immobilization was confirmed by the uniform intensities over the whole chip, e.g. Figure 7.9.

After use, the PDMS slab was washed with: sparkleen solution, LiquiNOX, DI water, 95% ethanol, and DI water, and dried with air. Then the slab was reused and sealed back onto the glass slide with the channels running vertically.

b) PCR strand Immobilization

The 16 channel PDMS slab was washed with: sparkleen solution, LiquiNOX, DI water, 95% ethanol, and DI water, dried with air and sealed to the aldehyde glass slide such that the channels run horizontally. A solution containing 30 ng/μL of asymmetric PCR products (PCRb, PCRc, with P7 primer in excess), 2× PBS and 100 mM $NaCNBH_3$ was applied to the channels and suction was used to pull the solution into the channels. The solution was incubated at room temperature for several hours and then removed using suction. The channels were then washed with 1× PBS followed by the removal of the PDMS slab from the glass slide. The glass slide was submerged in $NaBH_4$ solution (~2.5 mg/mL $NaBH_4$, 0.75× PBS, 23.75% ethanol) for 15 min, washed with 1× PBS and dried with nitrogen gas. The PDMS slab was washed with: sparkleen solution, LiquiNOX, DI water, 95% ethanol, and DI water, dried with air; the slab was reused and sealed back onto the glass slide with the channels running vertically.

ii) Hybridization

In the second step of NAT experiments, we performed either SF or dynamic hybridization. In the former case, we introduced sample strands by suction and allowed time for hybridization under the SF condition. In the latter case, sample introduction and hybridization were accomplished dynamically using centrifugal liquid pumping.

a) Stop Flow Hybridization

The SF method was used for probe-target (Method 1), target-probe (Method 2) and three-strand (Method 3) hybridizations.

1.0 μL of solution containing 1× SSC, 0.1% SDS and 25 nM of oligonucleotide or 40 ng/μL of PCR product was applied to the channel wells and suction was used to draw the solution into the channels. It was incubated at rt for 1 h and then excess solution was removed by suction.

b) Dynamic Centrifugal Hybridization

This dynamic flow method was used for probe-target hybridizations; whereas the first step of probe immobilization was conducted by stop flow (see section i).

The channels were washed with 0.5 μL of 1× SSC, 0.1% SDS solution. The chip was placed in the centrifugal chip holder and adhesive tape was used to hold the chip in place (Figure 2.6a). Next, 1.0 μL of solution containing 1× SSC, 0.1% SDS and 25 nM of oligonucleotide or 40 ng/μL of PCR product was applied to the channel wells and a small amount of suction was used to bring the solution to the first kink in the channels. The holder was then placed on the rotating platform and spun (Figure 2.6b). Afterwards, the

solution was removed from the channels with suction. The flow rates of 16 channels were studied and the comparison results were presented in Table 18.

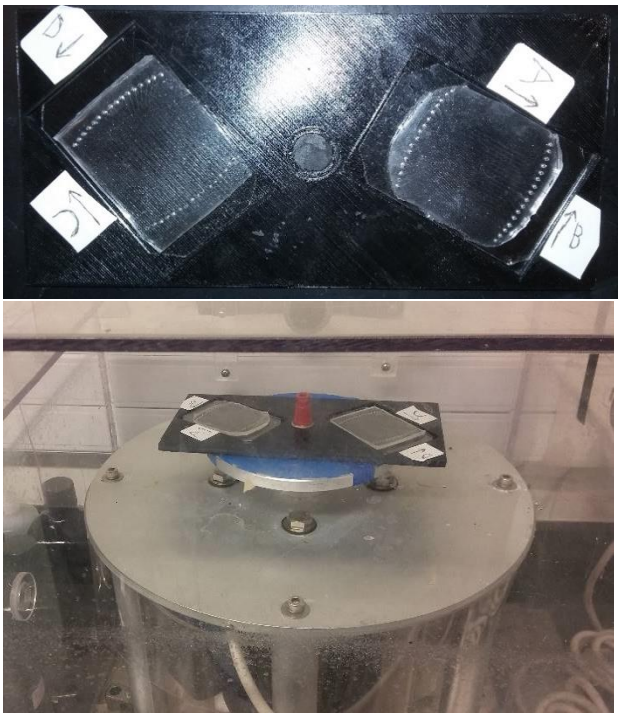


Figure 2.6: a) Photograph of the centrifugal chip holder with two chips on it. The four letters A, B, C and D represent the angles of the parallel portions of the channels to the radius which are 30°, 60°, 40° and 50°, respectively. Note that for direction B and D, the PDMS slabs have to be sealed perpendicular to the orientation presented in the figure. b) Photograph of the chip holder with two chips on it mounted on the rotating platform. Only one chip is used for an experiment at a time while the other chip functions as a counterweight.

iii) Fluorescence Labeling and Washing

After the hybridization step, 0.6 µL of SA-Cy5 solution (50 µg/mL in 1× PBS) was applied to the channels, drawn in by suction and incubated at rt for 15 min. The perfectly matched strand, which consisted of the biotin groups to react with streptavidin (SA), were thus labeled fluorescently by Cy5. The solution was then removed by suction and the channels were washed with 0.6 µL of 1× PBS/Tween. The PDMS slab was subsequently peeled off and the glass slide was washed with 1× PBS and dried with nitrogen gas.

The glass slide was scanned in a BioRad ChemiDoc Imager configured for Cy5 and automatic exposure duration. The scanned images were stored as .scn files. Image Lab was used to measure the fluorescence intensities of the microarray images from the files. The image on the scanner was inverted (to show high fluorescence as dark and low fluorescence as light). All the data files were copied onto a USB memory stick.

iv) Scanning and Image Analysis

The Image Lab software (freeware) was downloaded from the manufacturer's website (https://www.bio-rad.com/en-ca/product/image-lab-software?ID=KRE6P5E8Z accessed on Jun 30, 2021) and installed. This program was used to open the .scn file obtained from the scanner computer. (Note: this software has been made only for gel analysis, so the

following is what has been done to use the software for analyzing arrays). This software allows the microarray image to be analyzed with spot intensity output to a spreadsheet file, see Figure 2.7. This procedure is presented with screenshots in the Appendix A.

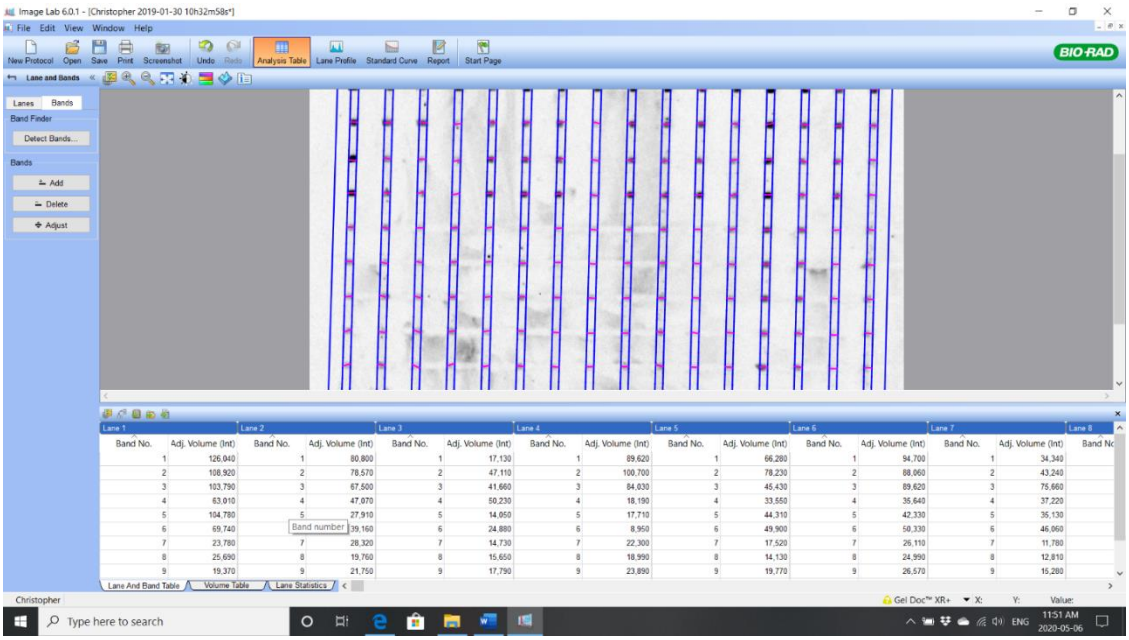


Figure 2.7: Image analysis of spots in 16 lanes and data output to a spreadsheet file (same figure as step 14 in Appendix A).

Each duplex hybridization was done in triplicates unless stated otherwise in the figure caption. The intensities shown in the figures are the averages of the replicates and their error bars are the standard deviations of the replicates and are related to the confident intervals for statistical significance.

Chapter 3 Different Strategies for Nucleic Acid Tests Using PCR Amplified Strands

We have examined three different nucleic acid test methods based on DNA hybridization: a) probe-target method (Method 1), b) target-probe method (Method 2), c) three-strand complex method (Method 3). In addition, hybridization has been conducted using either the SF or centrifugal dynamic condition.

a) Immobilized Probe-Target Hybridization (Method 1)

The conventional way of performing solid-phase hybridization is to immobilize probes on the solid surface and hybridize to them a labeled target (PCR strand) (Figure 3.1, also Figure 2.5a).

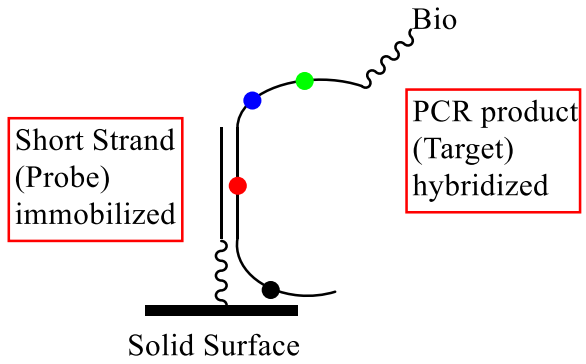


Figure 3.1: The probe-target method (conventional Method 1)-immobilized probe hybridized to a PCR strand labeled with biotin. The black, red, blue, and green circles on the PCR strand represent the Int, N1, N2 and N3 hybridization regions, respectively.

i) Utilization of the N1 and N2 Sites

Initially, amine probes were designed to hybridize at the N1 and N2 sites to biotin functionalized PCR products (targets). The probes used at N1 are A-N1G and A-N1Q; the probes used at N2 are A-N2G and A-N2Q. The fluorescence intensities obtained for these four probes are normalized with those of the non-specific probe Int, and so the results shown in Figure 3.2 are relative fluorescence intensities.

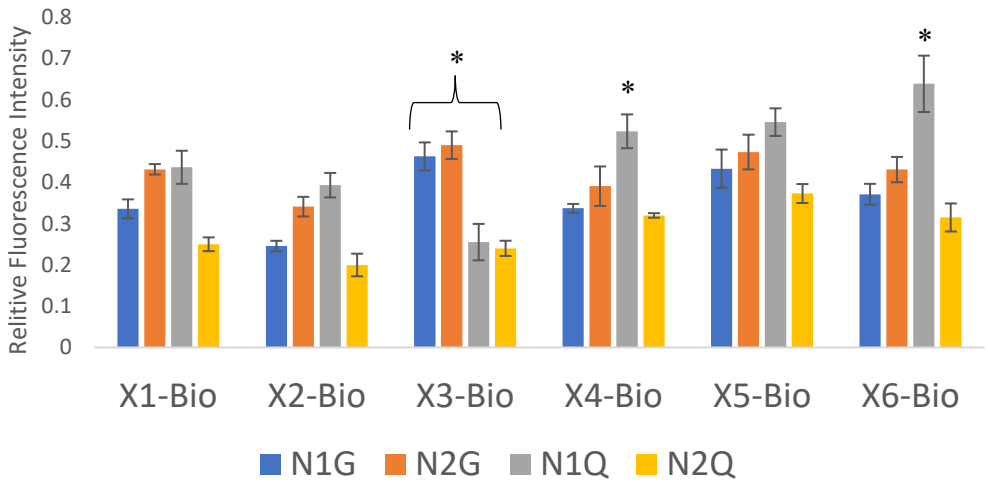


Figure 3.2: The relative intensities of biotin-labelled PCR products (Asymmetric, 226 nt) of the six ginseng powder samples hybridizing to immobilized A-N1G, A-N2G, A-N1Q and A-N2Q (Method 1). The relative intensities have been normalized with Int selected in the consensus region as an internal standard (see Figure 2.3). Raw image data can be found in Appendix C (Figure 7.1).

The results shown in Figure 3.2 indicate that in X3, both N1G and N2G have greater fluorescence than N1Q and N2Q (denoted by the asterisk); this led us to believe the powder sample X3 contained *P. ginseng*. In addition, X4 and X6 appears to contain *P. quinquefolius* since the fluorescence obtained for N1Q, if not N2Q, is significantly greater than that for N1G and N2G.

While the N2G and N1Q had resulted in strong fluorescence intensity, N2Q did not. To explain why the N2Q intensity was low for X4 and X6, the sequences of the complementary and mismatching probe-PCR product duplexes were examined *in silico* (DINAMelt and Mfold) and it was noticed that N2Q formed A-T bp with its respective complementary target and G-T bp with its mismatched target (shown in Table 7, and Appendix G). Since the G-T bp is partially stabilizing,[110] this mismatched duplex of N2Q will form more readily compared to a conventional mismatch (e.g. C-A bp) which is what is formed in the N2G and N1Q mismatches. This is further supported by both the *in silico* and experimental Tm values of the four probes (N1G, N2G, N1Q, N2Q) hybridizing to either a perfectly complementary strand or 1 bp mismatching strand (shown in the Appendix B).

While X3 and X4/X6 appear to contain pure *P. ginseng* and *P. quinquefolius*, respectively, this may not be the case for X1, X2 and X5. Figure 3.1 shows that X1, X2 and X5 could be a mixture of *P. ginseng* and *P. quinquefolius* or their hybrid since the N2G and N1Q intensities obtained for these three powdery samples are similar. More findings for X1, X2 and X5 are provided in subsequent chapters.

ii) Utilizations of Shorter Probes (N1 and N2) and the N3 Site

Several ginseng samples (powder and root) were submitted for Sanger sequencing by Genewiz. Only two samples gave definitive results; X3 (powder) had the sequence

consistent with *P. ginseng* and AmG2 (root) was identified as *P. quinquefolius*, see Appendix E. However, when the root sample AmG2 was analyzed by the same procedure as in section 3a(i): Utilization of the N1 and N2 Sites no identification as *P. quinquefolius* was observed with the N1Q probe (Figure 3.3), which is a discrepancy from the results shown in Figure 3.1. Two different issues were considered for such a discrepancy. The first issue is that the 30+ base N1Q probes might be too long to have sufficient differentiation given the fact that there is only one SNP site compared to 29 matched bp. Traditionally, the probes designed for microarray experiments range from 20-25 bases;[19,21,42] however, since the SNP sites of the DS gene are A-T rich, the probes were made longer than 25 bases in order to have a high enough Tm value for effective hybridization. This issue of long probe length might be a problem since the case of 1 in 30 bases mismatching would be less destabilized, and less differentiated compared to 1 in 20 bases. Thus, the shorter 20 base versions of N2G and N1Q were ordered (tN2G and tN1Q, respectively) and tested. A possible problem with a shorter probe (20 bases) could be a lower Tm and hence lower signal, as compared to a longer probe (30 bases).

The second issue for the use of the N1Q probe is that there is potential to have “slipping” occur in which a misalignment of the probe and target is involved. This "slipping" is especially problematic if within the probe sequence you have a 3+ base sequence involving the SNP base repeated elsewhere in the probe.[36] One example of this can be seen highlighted in grey in the sequence of the N1Q: C TAAAAAAAAA GTATTTCTCA TCTAAATTTT GAA (the N1 site is underlined). This "slipping" issue would weaken the effectiveness of hybridization using N1Q. Since the *P. quinquefolius* specific probe was needed, a new pair of probes (N3Q and LN3Q) was designed, i.e. LN3Q: AAAGTCATAT AATTTAGAAC GGAGGGAGT (N3Q is in red), with the N3 site underlined. We note that the three-base sequences involving the SNP base 'A' (i.e. TTA, TAG and AGA) are not repeated anywhere else within the new probe sequence.

The results that show the use of short probe (tN1Q), long probes (N1Q, N2G), and probes with the N3 site (short N3Q, long LN3Q) are depicted in Figure 3.2.

As shown in Figure 3.3, N2G hybridizes well only with X3, and this is in agreement with the result as seen previously in Figure 3.2. Even though X3 generated a lower signal with tN1Q than with N1Q, the signal with tN1Q was not low enough in order to differentiate between X3 and AmG3. On the other hand, tN1Q did not produce high enough intensity when hybridized with AmG3, probably because of a shorter probe length and lower Tm value of tN1Q compared with N1Q.

In addition, it was observed that N3Q and LN3Q hybridize well to AmG3 as predicted by no "slipping" in the new probe. However, N3Q and LN3Q also hybridize well with X3, failing the differentiation capability by these new probes. This means that when N3Q and LN3Q are used in the probe-target Method 1, the hybridization process is highly favorable, and is thus not influenced much by the lengths of probes and the presence/absence of a mismatch.

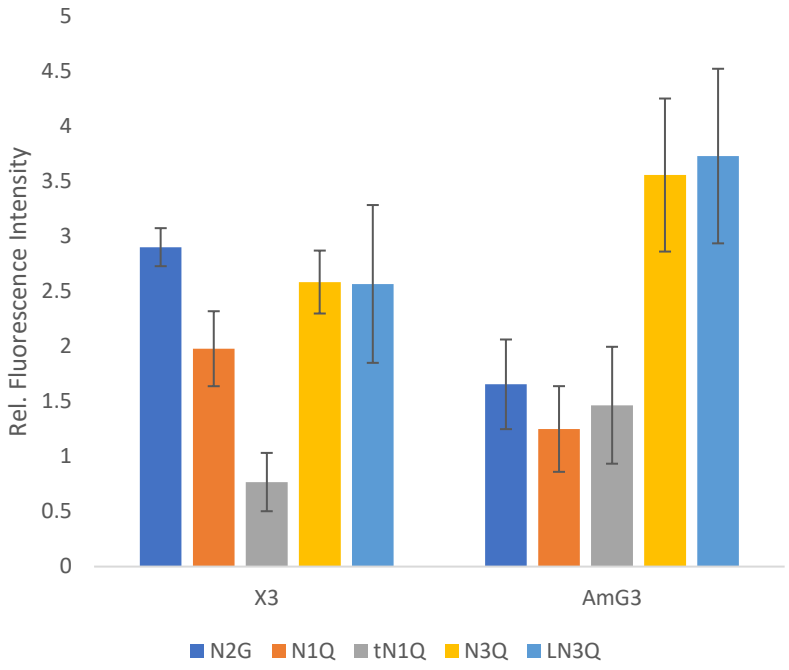


Figure 3.3: Fluorescence intensity of biotin-labeled X3 and AmG3 PCR strands (226 nt) hybridized to the immobilized probes: N2G, N1Q, tN1Q, N3Q, and LN3Q (Method 1). The relative intensities have been normalized with P8'. Raw image data can be found in Appendix C (Figure 7.3).

iii) Comparison of Three Lengths of PCR Strands

One more aspect of the probe-target method that was examined was the PCR product length. PCR products of three lengths (126, 226 and 694 nt) were produced. All three lengths of PCR strands were hybridized to the immobilized probes (Figure 3.4). It was observed that tN2G works better than N2G at identifying *P. ginseng* especially when the 126bp PCR product is used. The N3Q probe works better in identifying *P. quinquefolius* also when using this 126bp PCR product; the N3Q hybridizes non-specifically when using the medium 226bp PCR products. The long 694bp PCR product has low intensities thus indicating that it is not very useful for species identification, possibly due to secondary structures formed by intramolecular folding.

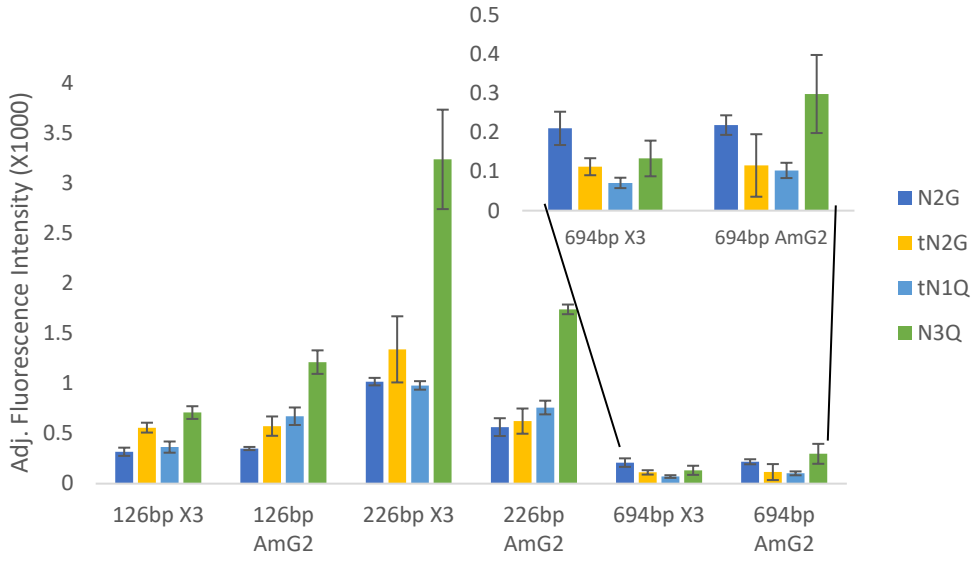


Figure 3.4: Fluorescence intensity of biotin-labeled X3 and AmG2 (126 nt, 226 nt, 694 nt) PCR strands hybridized to the immobilized probes: N2G, tN2G, tN1Q and N3Q (Method 1). This graph is normalized to factor in the differences in MW of the PCR strands. The inset shows the results in the expanded scale for the 694 nt PCR strand. Raw image data can be found in Appendix C (Figure 7.3)

From these results, we conclude that X3 can be differentiated well using the 226 nt PCR strands and N2G probe; AmG2 is differentiated well when 126 nt and N3Q are used.

In this section, we note that the conventional probe-target method, though worked well previously with samples of similar PCR strand length and SNP sites,[18,19,111] did not work well here for ginseng differentiation. Nevertheless, although not ideal, the probe-target method did allow us to differentiate X3 from AmG2 using the proper combination of PCR strand length and SNP site.

We then proceed to explore a new method, called target-probe hybridization, for ginseng species differentiation.

b) Immobilized Target-Probe Hybridization (Method 2)

We theorized that it could be easier for the short (biotin labeled) probe to hybridize to an immobilized long target (new target-probe Method 2 in Figure 3.5, also Figure 2.5a) than it would be for a labeled long target to hybridize to a short immobilized probe (conventional probe-target Method 1 in Figure 3.1) because there are no strands colliding with the solid surface in the new Method 2. Method 2 may also allow us in the future to conduct *in situ* bridge amplification[109] on the solid surface to increase the density of immobilized strands.

The first procedure to implement Method 2 was to prepare amine functionalized PCR products in the immobilization buffer. However, since it was not possible to achieve a high concentration (i.e. 25 μM) of the PCR target strand, no fluorescence signal was observed regardless of what probe concentration (i.e. up to 250 nM) was used in the hybridization step.

Therefore, we attempted to conduct optimization of DNA immobilization by performing an *in situ* reduction.

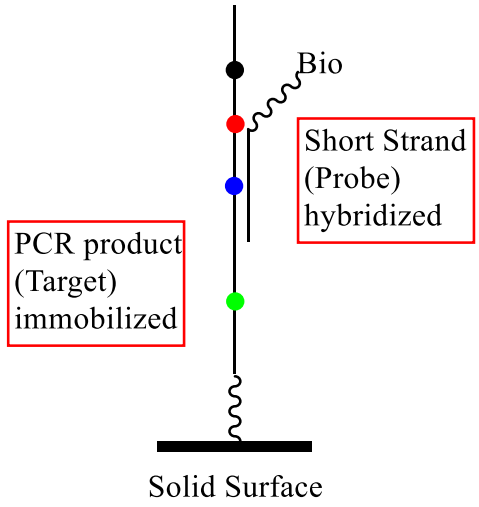


Figure 3.5: The target-probe method (Method 2)-immobilized PCR strand hybridized to a detector probe labeled with biotin. The black, red, blue, and green circles on the PCR strand represent the Int, N1, N2 and N3 hybridization regions, respectively.

i) Immobilization Time

Immobilization of DNA involves the reaction of the 5'-amine-conjugated DNA and the aldehyde containing solid surface to form the Schiff base. This base is chemically reduced (reaction a) in the subsequent step of the procedure using $NaBH_4$ (refer to section 2 f (i): Strand Immobilization). On the other hand, sodium cyanoborohydride ($NaCNBH_3$), which is a softer reducing agent compared to $NaBH_4$ will reduce the Schiff base without reducing the unreacted aldehydes. Therefore, this $NaCNBH_3$ reduction (reaction b) allows the amine containing reagent (DNA), aldehyde containing reagent (solid surface) and $NaCNBH_3$ to be combined into a single reaction resulting in an *in situ* reduction which pushes the equilibrium to the product side to generate more Schiff base. In addition, this reaction b is done at either neutral or mildly acidic pH instead of basic pH and $NaCNBH_3$ is stable in water unlike $NaBH_4$ (2× PBS was thus chosen for this reaction b). This is a common technique for conjugating proteins/peptides to aldehyde functionalized auxiliaries.[112] The benefits of selectivity and stability of $NaCNBH_3$ come at the cost of reactivity meaning that the time of reaction b will be longer than that of $NaBH_4$ (usually 1 h) which needs to be optimized to this solid-phase reaction. Furthermore, even though the immobilized PCR product has its Schiff base linkage reduced, there are still unreacted aldehydes, which are left over on the glass surface, to be removed thus requiring the $NaBH_4$ step to be performed after the immobilization.

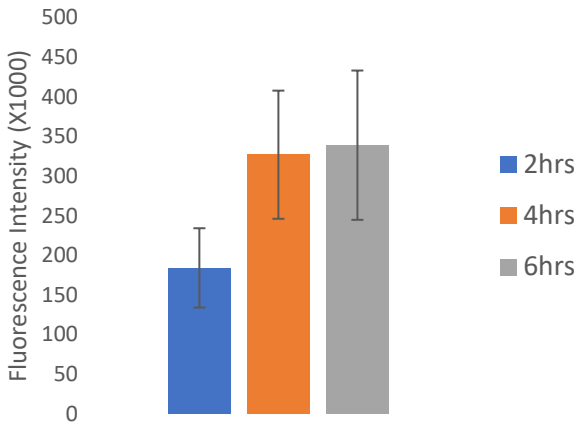


Figure 3.6: Immobilization of 226 nt AmG2 PCR strand to the solid surface using $NaCNBH_3$ and incubating it for 2, 4 and 6 hrs.: B-N2Q was used for detection. Raw image data can be found in Appendix C (Figure 7.4).

As seen in Figure 3.6, a duration of four hours is the optimal reaction time for the immobilization of the PCR strands and so 4 h was selected for all subsequent experiments.

ii) Immobilization of oligonucleotides vs. PCR Strand

Since this PCR immobilization method was new (reaction b) for use with the microfluidic chip, there was a need for reaction b to also work for oligo-immobilization to have a proper positive control. Oligonucleotide strands A-Gin and A-Quin were chosen, and 100nM in the same PBS/$NaCNBH_3$ buffer was used since this was of a similar molar concentration as the PCR strands. After strand immobilization and the standard 1 h SF hybridization method was used, Gin had strong fluorescence regardless of which of the two probes (N2G and N1Q) was used (Figure 3.7a), which was expected for N2G but not for N1Q. When the same reaction were conducted using dynamic hybridization, the expected pattern was achieved with Gin hybridizing preferentially to N2G and Quin hybridizing preferentially to N1Q (Figure 3.7b). The development of the dynamic hybridization methodology based on centrifugal pumping will be discussed in more detail in the section 3d: Dynamic Hybridization using Centrifugal Chip Holder. What is also evident in Figure 3.7 is that the fluorescence intensity with N1Q is higher than that for N2G regardless of the hybridization method. This can be explained by the fact that the N1 hybridization site on the Gin/Quin is further away from the solid surface than the N2 site (see Figure 3.5), therefore, the N1Q probe has less steric hindrance accessing its hybridization region on the immobilized strands.

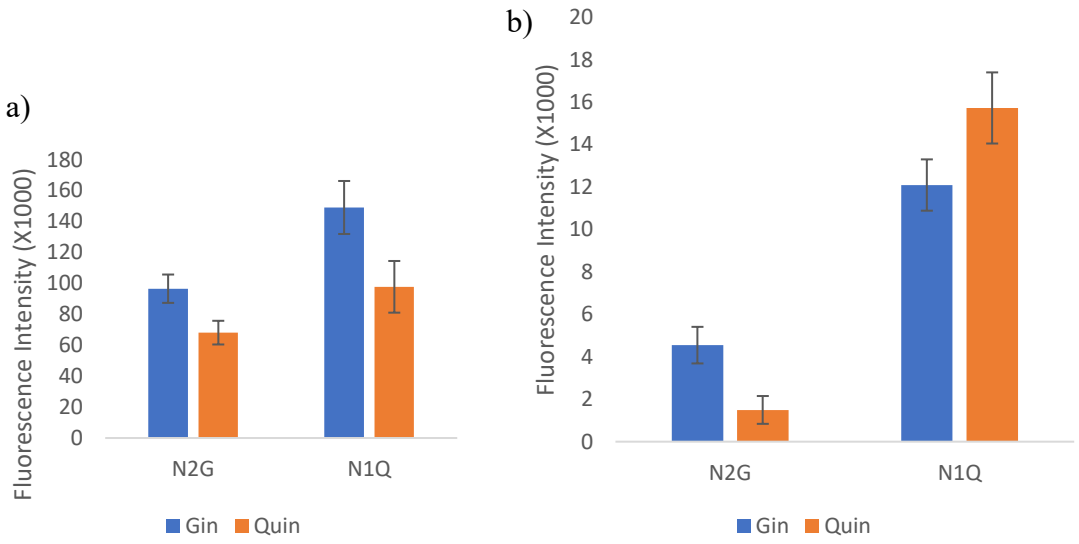


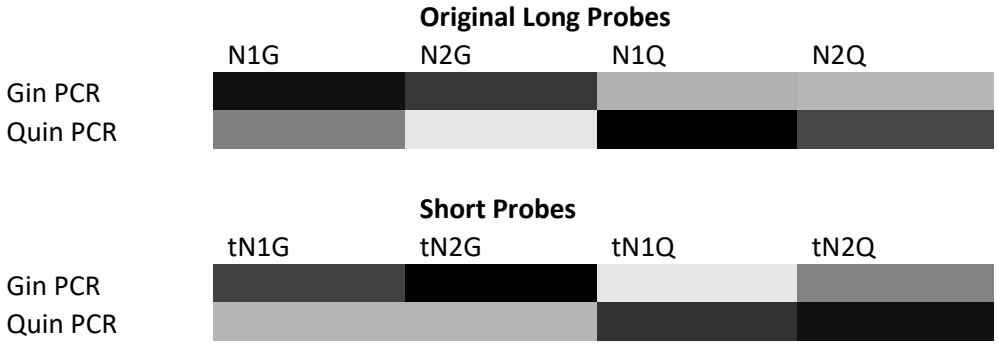
Figure 3.7: Hybridization of immobilized A-Gin and A-Quin to B-N2G and B-N1Q (Method 2), a) SF hybridization, b) dynamic hybridization (1 injection of 1.0 µL, 5 spins). Raw image data can be found in Appendix C (Figure 7.5).

After the study of immobilized oligonucleotide strands was completed, X3 and AmG2 PCR strands were next immobilized and hybridized to N2G, N1Q detection probes. What was observed was that when SF hybridization was used, it was good that X3 had a strong fluorescent signal only when hybridized with N2G (Figure 3.9a); but even so there was no differentiation for the AmG2 like what we observed for the oligonucleotide Quin strands (Figure 3.7). When the experiments were performed using the dynamic hybridization method, it was good that the AmG2 had a strong fluorescent signal only when hybridized with N1Q; but there was no strong X3-N2G signal to differentiate X3 from AmG2 (Figure 3.9b).

These results cannot be explained by the thermodynamic predictions, see the heat map shown in Table 7, and hybridization kinetics must be considered. Hybridization rate constants can be experimentally determined by surface plasmon resonance (SPR) as previously reported.[19] On the other hand, the hybridization kinetics has been theoretically and experimentally studied by the Rice University's group[36] and a software tool is recently available from them (nablab.rice.edu/nabtools/kinetics). From this tool (using salinity of 0.15 M Na^+ and temperature of 22°C), the hybridization rate constants predicted for the sequences (N2G and N1Q) are calculated to be 2.04 x10^6 and 2.23 x10^6 $M^{-1}s^{-1}$, respectively (see Appendix D, Table for full list of probe hybridization rate constants). Therefore, we speculated that AmG2 hybridizes to the N1Q probes more quickly than X3 and as a result, N1Q requires a shorter reaction time (achieved in dynamic hybridization) in order to achieve differentiation between perfectly complementary hybridization (AmG2) and mismatch hybridization (X3). On the other hand, though N2G forms a more thermodynamically stable duplex (more negative ΔG) with X3 than with AmG2, we speculated they both have slower hybridization rates to

N2G than to N1Q and so N2G require longer reaction times in SF hybridization (than dynamic hybridization) to achieve sufficient signal and proper differentiation.

Table 7: Heat map showing relative binding (ΔG) based on DINAMelt thermodynamics of the originally long probes (N1G, N2G, N1Q and N2Q) and short counterparts with the 226 nt PCR strands. The darker the rectangle is in the heat map, the more negative is the ΔG value of hybridization.



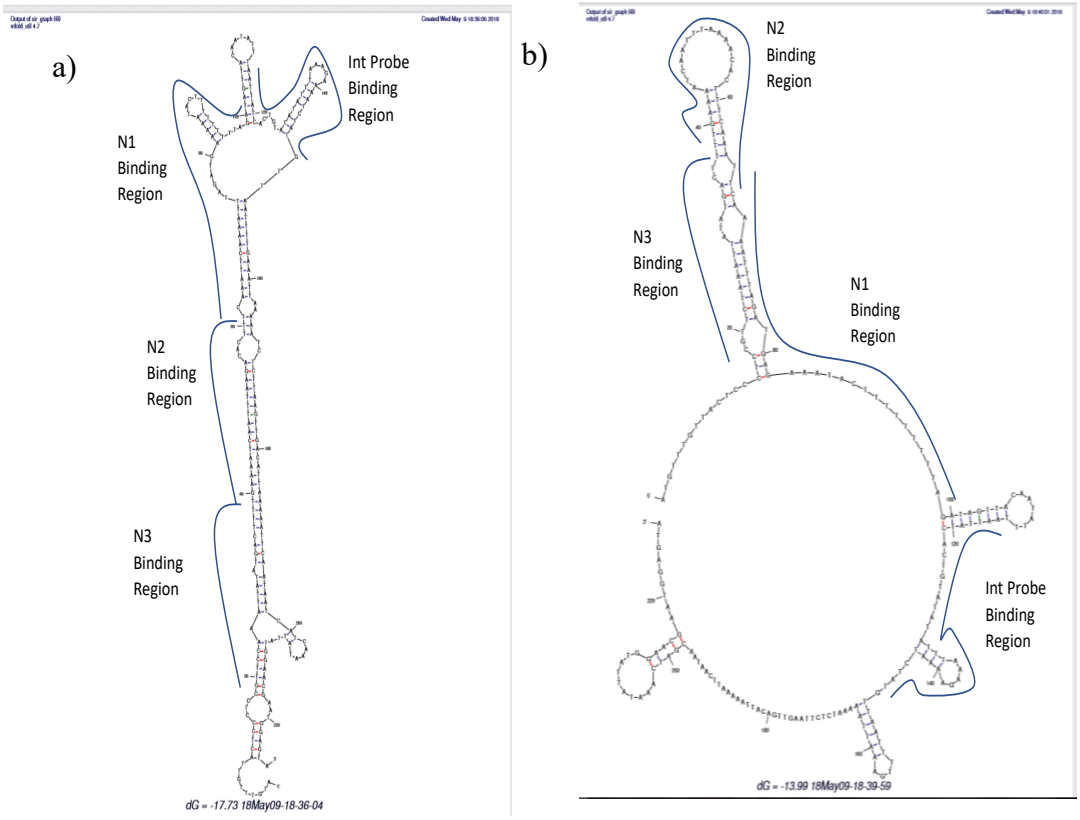


Figure 3.8: Mfold-predicted intramolecular folding of the 226 nt antisense strands of a) *P. ginseng* and b) *P. quinquefolius*. Both structures were predicted using 0.15 M Na^+ and 22°C.

This slow hybridization rate of X3 to N2G can be explained by secondary structures of *P. ginseng* and *P. quinquefolius* (Figure 3.8a). The entire *P. ginseng* antisense strand (226 nt) folds into a hairpin structure meaning that in order for it to hybridize to a probe, the

hairpin has to partially denature. This results in the hybridization reaction requiring more time while having higher specificity. The *P. quinquefolius* on the other hand has a region of exposed bases at both the N1 and N2 hybridizing regions (Figure 3.8b) which results in the hybridization reaction being much faster and losing specificity very quickly. Looking at the folding of the *P. quinquefolius* PCR strand, it can be seen that the N3 hybridizing region is part of a large hairpin structure similar to the N2 region of the *P. ginseng*. Thus, if the above analysis is correct, then the N3Q probe should have good differentiation for *P. ginseng* and *P. quinquefolius*, and for X3 and AmG2.

Hence, for the new target-probe Method 2, N2G, but not N1Q, can properly differentiate X3 from AmG2 using SF hybridization. However, dynamic hybridization should be selected for their differentiation based on N1Q.

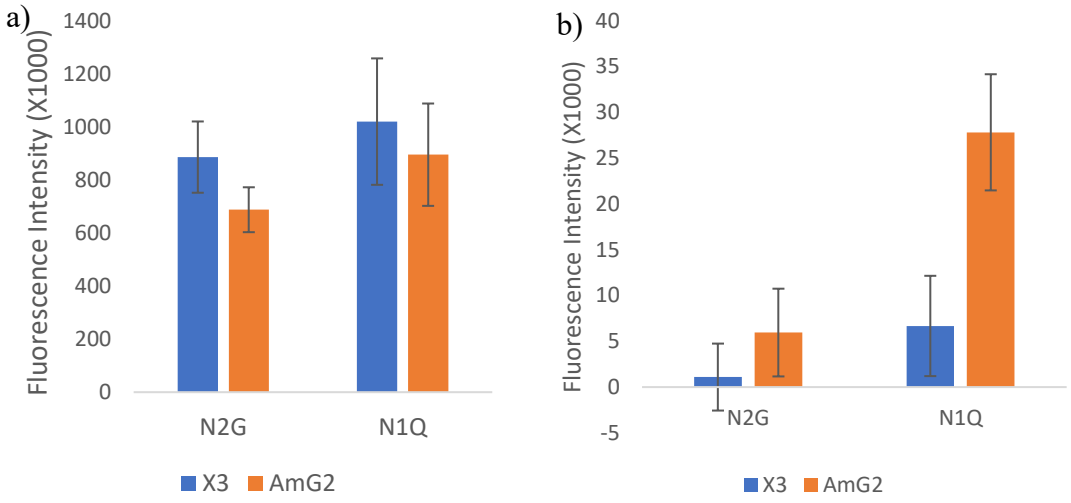


Figure 3.9: Hybridizations of immobilized 226 nt X3 and AmG2 PCR strands to B-N1G, B-N2G, B-N1Q and B-N2Q detection probes (Method 2): a) SF hybridization (1 h) b) Centrifugal (dynamic) hybridization using a single 1.0 µL injection. Raw image data can be found in Appendix C (Figure 7.5).

iii) Comparison of the Lengths of Probes, and N3 SNP Site

The next set of conditions that was examined was the different lengths of probes used (i.e. tN2G vs. N2G and tN1Q vs. N1Q). The use of short probes reduces the fluorescent intensity without improvement in the differentiation as shown in Figure 3.10. We speculate the lower intensities are caused by the lower Tm values of the 20 nt probes as compared to those of the 30 nt probes. These results of short probes used in Method 2 was consistent with those of Method 1.

Another condition to examine is the N3 site that did not provide differentiation using Method 1. In Method 2, N3Q does show differentiation favouring hybridization to its perfectly complementary AmG2 over mismatching X3, like the case with the 226 nt PCR strand as shown in Figure 3.10. There are two factors that have contributed to the successful differentiation ability. The first factor is that the N3 hybridization site on the immobilized target strand (as compared with N1 and N2) is the site closest to the solid surface (Figure 3.5) thus making N3 less accessible to the detection probes in solution.

This in turn makes the mismatch hybridization by X3 more difficult than the matched one by AmG2 to occur; this is consistent with the previously reported finding that a probe with a shorter tether produced better differentiation or specificity.[42] The second factor is that as was discussed in section 3aii), the N3Q probe does not have a three base sequence involving the SNP repeated elsewhere in the probe thus reducing the chance of 'slipping' or misalignment.

Therefore, in this new target-probe Method 2, we would continue to use the 30 nt long probe: N2G to differentiate X3 but switch to N3Q (from N1Q) in order to identify AmG2, all based on SF hybridization.

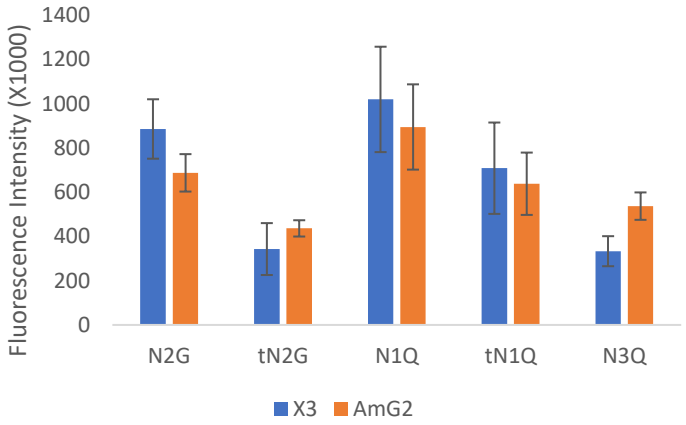


Figure 3.10: SF hybridization of immobilized 226 nt X3 and AmG2 to B-N2G, B-tN2G, B-N1Q, B-tN1Q and B-N3Q (Method 2). The hybridization was performed using the 1 h SF method. Raw image data can be found in Appendix C (Figure 7.6).

iv) Long (226bps) vs. Short (126bps) Products

Another aspect of studying this Method 2 (using immobilized PCR strands) was the effect of the PCR strand length on hybridization signals. Two different PCR strand lengths were compared: 226 nt and 126 nt. Although the motivation of the comparison was to anticipate more effective hybridization from a shorter PCR strand, possibly with less hairpins or secondary structures, we did not see such a difference. What is seen with X3 (226 nt) and X3' (126 nt) is that regardless of what PCR strand length is used, the hybridization pattern to the probes remain the same (

Figure 3.11a). For the AmG2 PCR strands, the lower signal of the 126 nt PCR product (AmG2') than the 226 nt PCR strand (AmG2) is possibly caused by a lower Tm value of a shorter strand. However, as seen in

Figure 3.11b this has not uniformly affected the probe hybridization compared to the 226 nt product (AmG2) as indicated by the ratios of relative fluorescence placed above the bars.

Therefore, it is determined that, in the differentiation between *P. ginseng* and P. *quinquefolius*, the long 226 nt PCR strand would be used; N2G had produced a higher

signal for X3 than AmG2, and N3Q would differentiate AmG2 over X3 (as seen from the chart in

Figure 3.11).

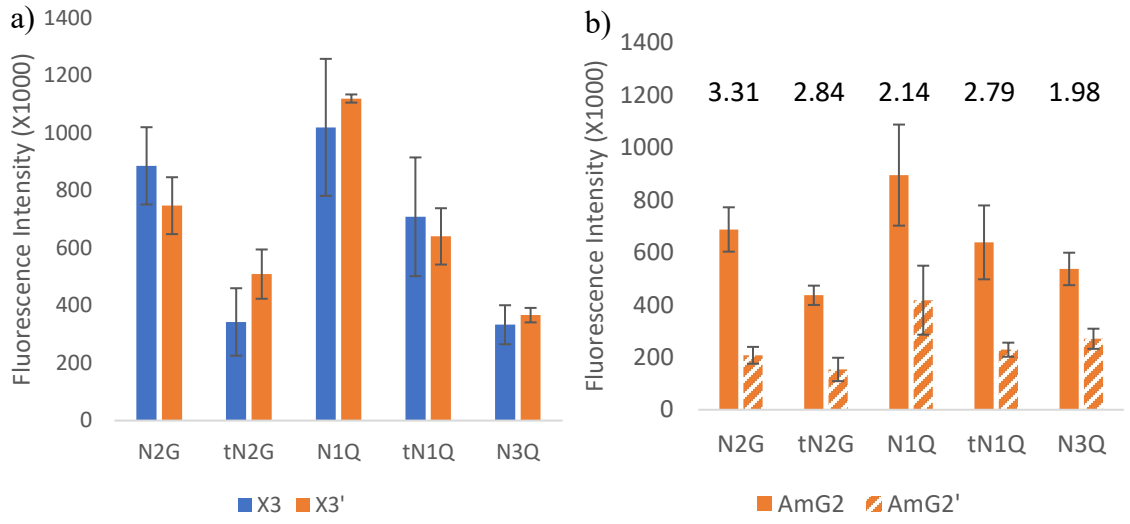


Figure 3.11: Comparison of the effect of the length of PCR strands on their hybridization to the probes B-N2G, B-tN2G, B-N1Q, B-tN1Q and B-N3Q (Method 2). All hybridization was done using the 1 hr SF method. a) X3 and X3' are the 226 nt and 126 nt PCR strands, respectively. b) AmG2 and AmG2' are the 226 nt and 126 nt PCR strands, respectively. The intensities of the 226 nt PCR strands in a) and b) were used to construct the graph in Figure 3.10. Raw image data can be found in Appendix C (Figure 7.6).

c) Immobilized Capture Strand-Target-Probe Hybridization

Another method that was investigated would be considered somewhat like a hybrid method between the Immobilized Probe-Target Hybridization (Method 1) and Immobilized Target-Probe Hybridization (Method 2), but unlike either of them, three strands were involved in the hybridization (Method 3). In this third method (Method 3), an amine labeled oligonucleotide (PCR primer), which is the capture strand, is immobilized—using the 1 h conventional method (reaction a, 25µM concentration in immobilization buffer) rather than the 4 h procedure (reaction b), and then the PCR strand is hybridized to the immobilized capture strand (conserved region). However, unlike Method 1 where the PCR strand is biotin labeled, here in Method 3 there is no label on the PCR strand. Instead, biotin labeled detector probes are introduced (also in hybridization buffer) to hybridize to a SNP region on the PCR strand to form a three-strand DNA complex consisting of two oligos (primer and probe) hybridized to different regions of the one PCR strand (Figure 3.12).

There are two ways to immobilize the capture strand so that the orientations of the target (PCR strand) may differ. First, the capture strand was immobilized at 5' end so that the PCR strand was oriented away from the solid surface (Figure 3.12); second, the capture strand was immobilized at the 3' end and the PCR strand was close to the surface (Figure

3.15). The two orientations of the PCR strand produced different results, as described in detail as follows.

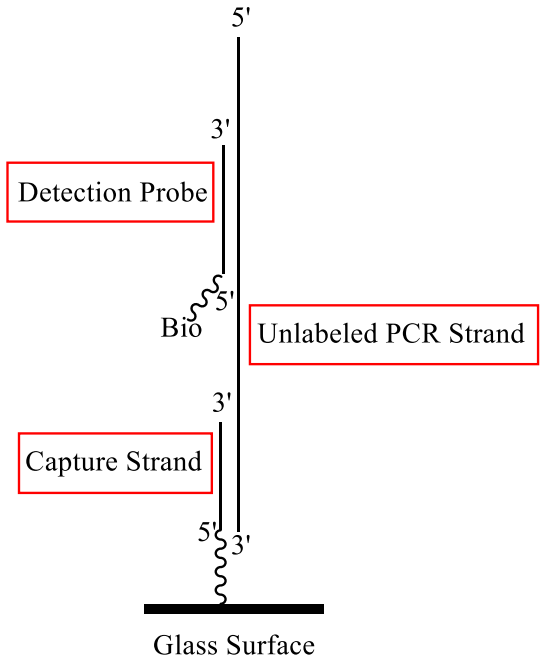


Figure 3.12: Schematic of a three-strand complex using a 5' amine capture strand, unlabeled PCR strand and a detector probe labeled with biotin (Method 3).

i) Using a 5'-NH_2 Capture Strand

When this three-strand complex method (Method 3) was used with a 5' amine capture strand, the correct differentiation was seen with the N3Q probe (Figure 3.13); this result is consistent with that obtained by Method 2. Undesirably, none of the N2G, tN2G, N1Q and tN1Q detection probes provide any differentiation.

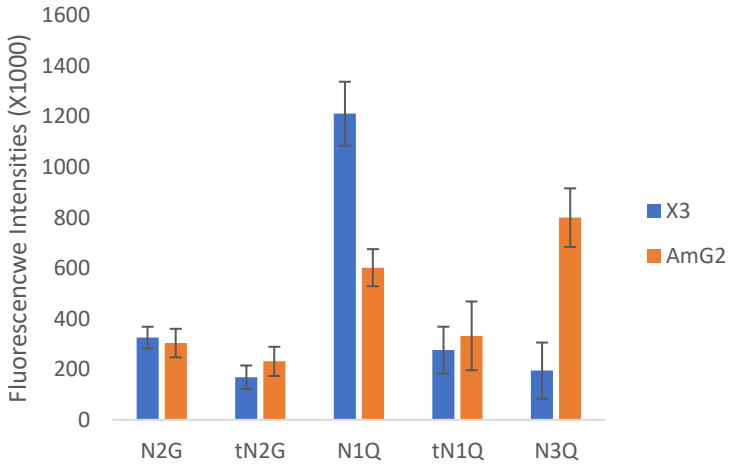


Figure 3.13: Three strand complexes made by hybridizing unlabeled 226 nt asymmetric X3 and AmG2 to immobilized 5'-A-P8' capture strand (Method 3). Then B-N2G, B-tN2G, B-N1Q, B-tN1Q, B-N3Q are hybridized to those captured

PCR strands (SF, 1 h). Each three-strand complex was done in six replicates. Raw image data can be found in Appendix C (Figure 7.7).

ii) Comparing PCR Strands of Different Lengths

Another aspect of Method 3 that was investigated was the PCR strand length. Three different PCR strand lengths (694 bp, 226 bp and 126 bp) were captured onto the solid glass surface. The trend that was observed was that as the PCR product length was shorter, the signal intensity (hybridized with N2G, N1Q) increases, but the differentiation ratio remains to be similar (Figure 3.14a). Two factors can be used to explain this observation. The first is simply that since 40ng/µL solutions of PCR products were used, this equates to a higher molar concentration for lower molecular weight (shorter) PCR products. This subsequently results in more captured PCR products for the N2G and N1Q to hybridize to, see Figure 3.14b when this factor is accounted for. The second factor is that the longer PCR products have more intramolecular folding which makes it more difficult for both the immobilized capture probe and detector probe to hybridize to the PCR product.

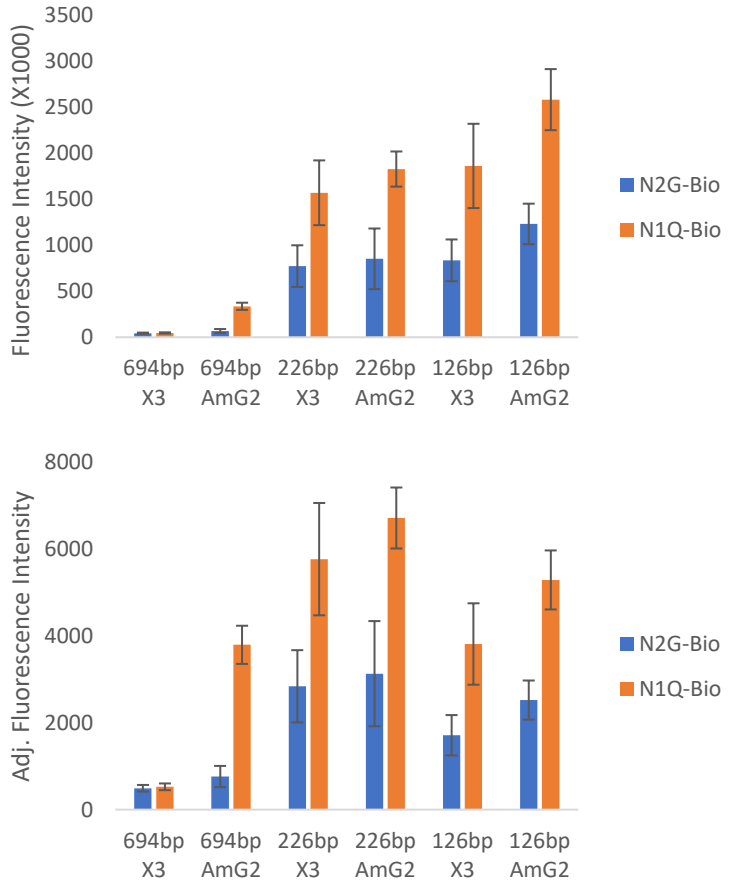


Figure 3.14: Three strand complexes made by hybridizing unlabeled X3 and AmG2 (40ng/µL) to immobilized 5'-A-P8' capture strand (Method 3). Then B-N2G, and B-N1Q are hybridized to those captured PCR products. Each three-strand complex was done in six replicates and both hybridization steps were done using SF hybridization. The top graph was constructed using the fluorescence intensities and the bottom graph was constructed from the quotient of those same intensities over the strand's concentration in nanomolar. Raw data can be found in Appendix C (Figure 7.8).

iii) Using a 3'-NH_2 Capture Strand

It was examined what effect changing the position of the amine from 5' to 3'on the capture strand would have. As seen in Figure 3.15, flipping the orientation of the capture probe causes the PCR strand to face down into the solid support which is the region where the biotin labeled probe needs to hybridize to.

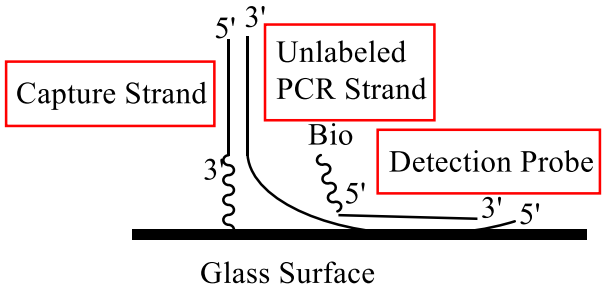


Figure 3.15: Schematic of a three-strand complex using a 3' amine capture strand (primer), unlabeled PCR strand and a detector probe labeled with biotin (Method 3).

Similar to the system using the 5' amine counterpart, the 3'-amine capture strand allowed AmG2 hybridizing more with N3Q compared to X3 (Figure 3.16a). More importantly, the system using the 3' amine shows the notable difference that there is correct differentiation of the two PCR strands using N2G. In both systems, the N1Q hybridizes undesirably well with X3 despite the mismatched base pairing. However, when this three-strand complex was formed using dynamic hybridization (for the biotin labeled probe hybridization step only) there was preferential hybridization of the N1Q and tN1Q for AmG2 over X3 (Figure 3.16b). This is a similar pattern to what was observed in Method 2 where the N1Q hybridization to a target is a kinetically driven process while N2G hybridization has thermodynamics playing a bigger role.

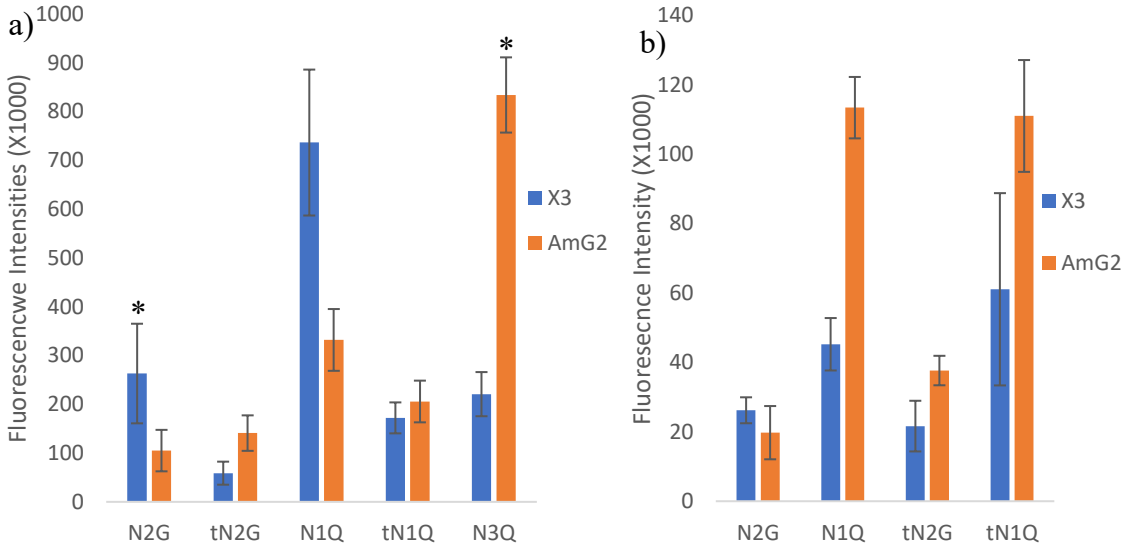


Figure 3.16: Three-strand complexes made by hybridizing unlabeled 226 nt X3 and AmG2 to immobilized 3'-A-P8' capture strand as the capture strand (Method 3). Then B-N2G, B-tN2G, B-N1Q, B-tN1Q, B-N3Q are hybridized to

those captured PCR strands. (a) Both hybridization steps were performed using SF hybridization b) The PCR strand hybridization was performed using SF hybridization and the probe hybridization was performed using dynamic hybridization. All three-strand complex were made in six replicates. Raw image data can be found in Appendix C (Figure 7.7).

In conclusion, for the three-strand method (Method 3), the capture strand is selected to be immobilized at the 3'-end, and N2G and N3Q are the detection probes of choice to differentiate X3 from AmG2.

d) Dynamic Hybridization using Centrifugal Chip Holder

In preceding sections, it has been concluded that dynamic hybridization differentiates AmG2 from X3 (using N1Q) better than the SF hybridization method. In this section, we describe the details about how dynamic hybridization is implemented by centrifugal liquid pumping.

The centrifugal chip holder (Figure 2.6a) has been designed to allow the conventional 75 mm X 50 mm chips to be used in experiments where centrifugal force, instead of suction, is used to push the solution through the channels. When the holder was tested initially using blue dye (to better monitor the fluid flow), it was realized that all four directions (labeled A, B, C and D as shown in Figure 2.6a) have some channels that do not flow at all (data shown in Appendix C Figure 7.10 and Appendix D Table 18). The reason for this was because the kinks at the ends of the channels have made the channel angles at the wells greater than 90°C relative to the radius resulting in no centrifugal force being applied to the fluid. However, since there were only a few channels on one side on the chip that did not flow, there was a way to resolve this problem. Instead of injecting DNA solution into all 16 channels, the solution was injected into eight channels (1 to 8 in Figure 3.17) on the side of the chip that do all flow. Then the chip was flipped around, and the remaining eight channels (-8 to -1 in Figure 3.17) were utilized. This means that the directions of fluid flow on the right half (1 to 8) and left half (-8 to -1) of the chip are opposite (Figure 3.17). The background noise might be a result of the washing step and the occasional speck might come from a scattering center on the chip due to dust.

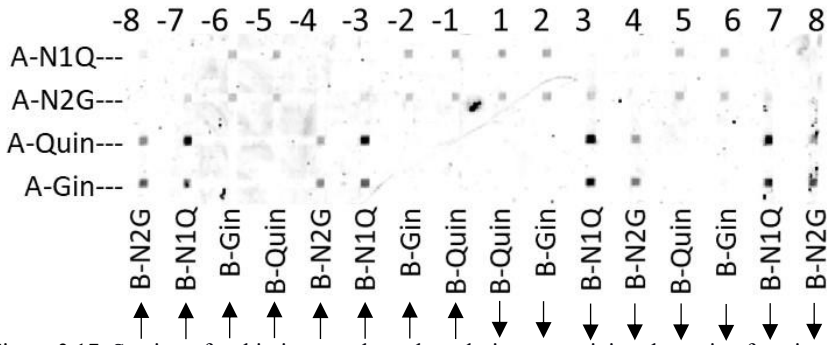


Figure 3.17: Section of a chip image where the solutions containing the amine functionalized oligonucleotides were applied to the horizontal channels and the hybridizing solution flowed through the vertical channels. The channels on the left half (-8 to -1) have the fluid flow from the bottom to the top and the channels on the right half (1 to 8) have the fluid flow from the top to the bottom. Fluid flow direction is also indicated by arrows in the vertical channels.

Another experiment was performed to examine the effect of using one injection of 1.0 μL of hybridizing solution verses 2 injections of the same solution, each 0.5μL (i.e. same total volume of 1.0 μL used). What was observed is that using the 2 injection method have stronger fluorescence signal compared to the single injection method (data available in Appendix D Figure 7.22). The reason for this is because during the centrifugation, the solution did not flow out of the channel but instead accumulated in the well (at the other end of the channel from the injection well). When the centrifugation was stopped, this solution back flowed back down the channel regardless of whether 1.0 μL or 0.5 μL of solution were used. Thus, using 2 injections of 0.5 μL allows the back flow to occur twice which gives the hybridizing strand more exposure time over the solid surface.

i) Probe-Target Hybridization (Method 1) vs. Target-Probe Hybridization (Method 2)

The centrifugal hybridization technique was performed on Method 2: target-probe hybridization (long oligonucleotide strand immobilized, short strand hybridized) and Method 1: probe-target hybridization (short oligonucleotide strand immobilized, long strand hybridizes). The target-probe Method 2 (Figure 3.18a) resulted in the expected pattern with the N2G hybridizing more to Gin compared to Quin, and N1Q likewise hybridizing better to Quin compared to Gin. The reason that the N1Q intensities are significantly greater than their N2G counterparts is because the N1 hybridization site is further from the solid surface compared to the N2 site and N1 is thus more accessible to the probes in the solution.

On the other hand, the probe-target Method 1 (Figure 3.18b) shows much lower fluorescence signals compared to the target-probe Method 2. Moreover, there is much less differentiation and there is no statistically significant difference, i.e. p value > 0.05, between the four hybridization pairs (see Appendix D Table 19). This indicates that in Method 1 when the hybridizing strands are long, the kinetics of the reactions become slower, and the presence of a mismatching base is not a hindrance to the process of hybridization.

Therefore, the target-probe Method 2 is found to benefit more from the use of dynamic hybridization than the probe-target Method 1.

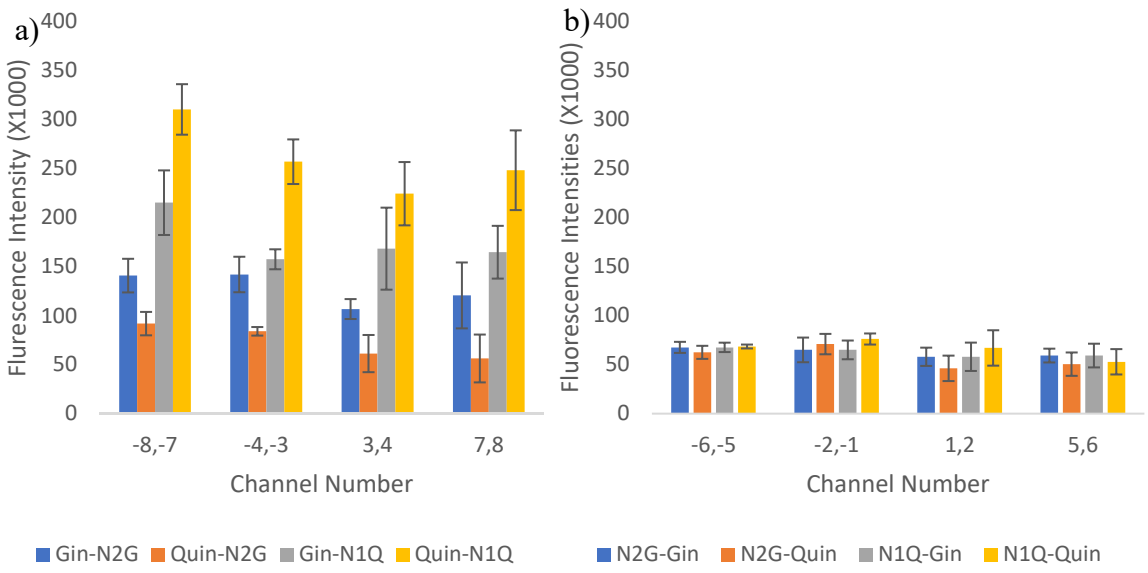


Figure 3.18: Hybridization of Gin/Quin targets to the N2G and N1Q probes using 2 injections of 0.5 µL. The channel numbering refers to the numbering used in Figure 3.17. Each hybridization was made in quadruplicates. a) hybridization of immobilized targets to biotin labeled probes (Method 2). b) Hybridization of immobilized probes to biotin labeled targets (Method 1). Raw image data can be found in Appendix C (Figure 7.9).

ii) Optimization of Probe-Target Method: Intensity Enhancement in Dynamic Hybridization

When dynamic hybridization was conducted with the probe-target method (Method 1), the fluorescence intensities of oligonucleotide targets were low. Therefore, we wanted to find ways to enhance the intensities of this method. PCR targets (even obtained in a greater amount by asymmetric PCR) were also tried using the probe-target Method 1 by dynamic hybridization, however, no quantifiable fluorescence signal was obtained. Some fluorescence could be seen at the intersections on the chip when the image contrast was adjusted, and so it was hypothesized that if the reaction residence time of PCR strands in the channels were increased, the fluorescent intensity would be enhanced to a quantifiable level. It was decided to implement this enhancement by exploiting the backflow phenomenon to increase the signals that was previously reported.[113] Each injection of 0.5 µL fluid was centrifuged and then allowed to backflow five times before being removed by suction. This method was first tested on oligonucleotides (Figure 3.19). The use of the five-spin method does enhance the fluorescence signal intensity compared to the use of a single spin. In terms of differentiation power of the target-probe Method 2, the increase in fluorescence impacts, but not to a significant extent, the differentiation by N2G (1.82 vs. 1.64 Gin/Quin for 1-spin vs. 5-spin, respectively). However, the 5-spin method does impact the differentiation power of the N1Q probe (1.52 vs. 1.32 Quin/Gin for 1-spin vs. 5-spin, respectively).

The five spin method has also enhanced the fluorescence intensity in the probe-target Method 1 compared to using a single spin (Figure 3.18b), however, the method has still shown no differentiation power as seen in Figure 3.19.

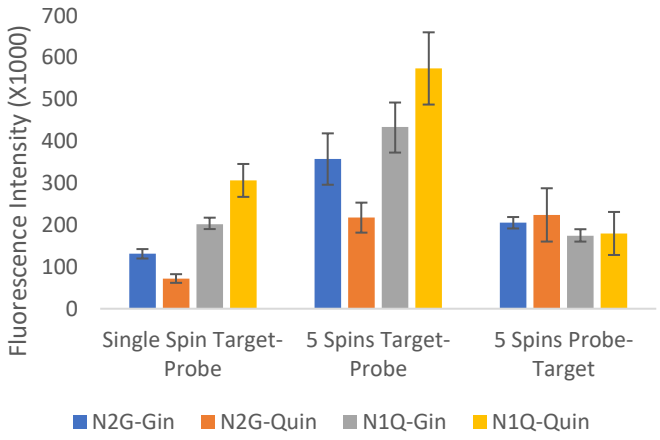


Figure 3.19: Comparison of the target-probe Method 2 using one-spin, target-probe Method 2 using five-spins and the probe-target Method 1 using five-spins. The signals for the probe-target Method 1 using one-spin were so low that they were not shown. Raw image data can be found in Appendix C (Figure 7.10).

iii) Probe-Target Hybridization of PCR Products (Method 1)

After enhancement of fluorescence intensity, the centrifugal hybridization method was next used to hybridize PCR products to immobilized probes (Method 1). The differentiation power was achieved using a denaturing agent, i.e. formamide. As seen in Figure 3.20a, when the standard hybridization buffer (1× SSC, 0.1% SDS) is used, X3 had the highest fluorescence intensity when hybridizes to N2G but AmG2 did not hybridize the strongest to the N1Q. When a small amount (2.5% v/v) of formamide was added to the hybridizing buffer (Figure 3.20b) it improved the hybridization of AmG2 to N1Q. None of the other probe-target duplexes were affected to a significant extent by the use of formamide. Further increasing the formamide concentration (5% and 10% v/v) did increase the fluorescence intensity further but the differentiation disappeared for both X3 and AmG2 (Figure 3.20c, d).

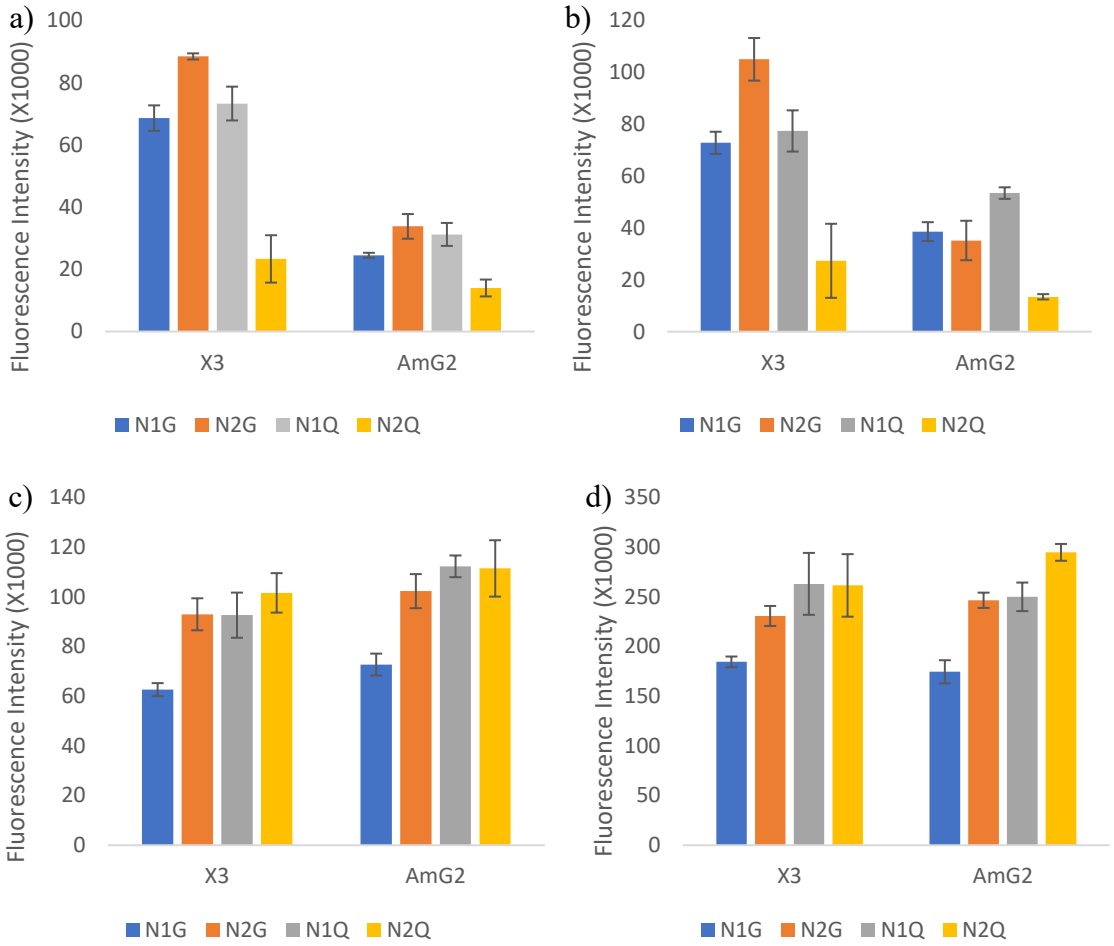


Figure 3.20: Centrifugal probe-target Method 1 hybridizing asymmetric biotin labeled 226 nt X3 and AmG2 to immobilized N2G and N1Q. The five-spin method was used with two injections of 0.5 µL each. Each injection was spun five times and the solutions was allowed to backflow after each spin. The hybridization buffer contained a) no additives, b) 2.5% formamide (v/v), c) 5.0% formamide (v/v) or d) 10% formamide (v/v). Raw image data can be found in Appendix C (Figure 7.11).

The results of using formamide showed that adding a small amount of denaturing agent helps improve the kinetics of hybridization probably by removing secondary structures in AmG2. Urea is another denaturing agent which is stronger than formamide, and urea is often used to prevent non-specific hybridization.[114] Hybridization buffer solutions containing 2, 4 and 6 molar of urea (data available in Appendix C, Figure 7.12) was seen to improve the fluorescence intensity, but urea does not improve differentiation similar to what was seen in Figure 3.20c and d using undesirably high amounts of formamide.

The probe-target Method 1 produced sufficient signal but inadequate differentiation at 5-spin. Nevertheless, this Method 1 showed an improvement in differentiation by the use of 2.5% formamide in the hybridization buffer.

iv) The Target-Probe Method 2 for Hybridization of PCR Strands

The dynamic centrifugal method has also been used in combination with the target-probe Method 2 for PCR strands. It has already been shown that the immobilized AmG2 resulted in a higher fluorescence intensity with N1Q than with N2G, because AmG2 hybridized more quickly with N1Q compared to N2G (Figure 3.9b). However, under these same conditions, the X3 did not hybridize to either probe to any notable degree; the SF method is thus required for significant fluorescence of X3 to be detected (Figure 3.9a). The dynamic hybridization is beneficial for the identification of AmG2 based on N1Q.

Dynamic hybridization has also been used in the three-strand method (Method 3) to enhance the differentiation of AmG2 based on N1Q (Figure 3.16b).

e) Comparison of the Three Different Nucleic Acid Test Methods

Three different hybridization strategies were examined: the probe-target method (Method 1), the target-probe method (Method 2), and the three-strand complex method (Method 3).

The N2G probe was able to differentiate between X3 and AmG2 using all three methods though Method 3 only has differentiation when the 3' amine capture strand was used. This probe was able to show the correct differentiation using either SF or dynamic hybridization when Method 1 was used. However, it was only able to show the correct differentiation using Method 2 and 3 when SF hybridization was used.

The N1Q probe was also able to differentiate AmG2 from X3 using all three methods but only when dynamic hybridization was performed. It should also be noted that Method 1 required 2.5% formamide to show differentiation.

The N3Q probe is the only probe to show the correct differentiation between AmG2 and X3 in Method 3 regardless of whether a 5' or 3' amine capture strand is used. In addition, it also shows the correct differentiation when Method 2 is used. The limitation of N3Q lies in that it does not achieve any differentiation when Method 1 is used.

The results comparing the different probes and methods are summarized in Table 8.

Table 8: The comparison of the three probes to each of the three methods as well as whether SF or dynamic (Dyn.) hybridization was used. "Yes" indicates that the method was able to successfully differentiate X3 from AmG2 and "No" indicate that the method could not differentiate the two samples.

Method	**N2G**	**N1Q**	**N3Q**	**Notes**
SF Method 1	Yes	No	No	
SF Method 2	Yes	No	Yes	
SF 5' Method 3	No	No	Yes	
SF 3' Method 3	Yes	No	Yes	
Dyn. Method 1	Yes	Yes	No data	N1Q requires 2.5% formamide
Dyn. Method 2	No	Yes	No data	N2G has no fluorescence signal
Dyn. 3' Method 3	No	Yes	No data	Dyn. method only used for second hybridization

The other aspect that should be considered is the duration and complexity of the experiments which are outlined in Table 9. Method 1 requires the least amount of time to perform (1 h for each step) and requires no special steps. On the other hand, Method 2 requires the reducing reagent to be present in the immobilization solution to allow for the *in situ* reduction which consequently results in this step being a 4 h reaction. This makes this hybridization method require the most time to perform out of the three methods. Lastly, Method 3 uses the same immobilization strategy as Method 1, however, it requires two consecutive hybridization steps (also in Table 8) instead of just one requiring 2 h to perform the two steps combined.

Table 9: List of immobilization and SF hybridization times of each of the three methods used. All hybridization times are reduced (to 10 min) when dynamic hybridization is performed.

Method	**Immobilization Time (h)**	**Hybridization Time (h)**	**Special Considerations**
Method 1	1	1	None, standard method
Method 2	4	1	Utilizes *in situ* strand immobilization and reduction
Method 3	1	2	Requires 2 sequential hybridization steps

f) Conclusions on the Nucleic Acid Tests of PCR Amplified Strands

Three different nucleic acid test methods based on DNA hybridization have been examined: probe-target method (Method 1), target-probe method (Method 2), and three-strand complex method (Method 3). Hybridization was conducted in either the SF or dynamic condition.

For the SF hybridization, the probe-target Method 1 has been demonstrated for the differentiation of X3 from AmG2 using N2G. However, the use of dynamic hybridization (5-spins) and formamide (2.5%) did not produce adequate differentiation by N1Q and N3Q probes. AuNP wash method was tried extensively, however, the method was never able to yield improved differentiation compared to a buffer wash, as shown in the Appendix D Figure 7.23. Therefore, other methods of nucleic acid tests were explored.

The target-probe Method 2 allows oligonucleotides to be differentiated by dynamic hybridization. However, Method 2 requires two different conditions to differentiate the PCR strands of X3 and AmG2: X3 requires the SF hybridization using the N2G probe while AmG2 requires dynamic hybridization with the N1Q probe. When switching to use the N3Q probe instead of the N1Q probe, AmG2 can also be differentiated from X3 using SF hybridization.

When the three-strand complex Method 3 is used, correct differentiation of both X3 and AmG2 is achieved. This is achieved by SF hybridization, only when the N2G and N3Q probes are used, respectively, and while using a 3' amine capture strand to immobilize the PCR strands (226 nt). Although dynamic hybridization does allow the correct differentiation using N1Q, this centrifugal method is not chosen because it comes at the cost of losing differentiation using the N2G probe.

Therefore, in proper differentiation of X3 from AmG2, the method of choice is the three-strand complex Method 3, with the 3' amine capture strand to immobilize the PCR strands (226 nt), using SF hybridization, and detection by the N2G and N3Q probes.

Based on the detailed studies of the three methods, several new design rules for nucleic acid test are developed and listed as follows:

1. Steric hindrance/SNP site location: When immobilizing the long (target) strand on a solid surface (Method 2), probe hybridization is fast when the probe is located far from the surface. In this way, there is low steric hindrance for the probe; the hybridization reaction between the target strand and the probe is fast and so the reaction is kinetically controlled. For instance, when the long target strand is immobilized in Method 2, N1 is farther from the surface than N2 and N3, and so dynamic hybridization benefits differentiation of target strands by N1Q.

On the other hand, there is no such benefit in Method 1 where the probe is immobilized and hence there is always steric hindrance for the hybridizing target strand.

2. Hybridization conditions (SF or dynamic): Probes with less secondary structure (e.g. N1Q) work better when using dynamic hybridization while probes with more folding (e.g. N2G) work better when using SF hybridization.

3. PCR strand length and secondary structures: The PCR strand length itself does not have a thermodynamic effect during hybridization. However, the length may have an effect on intramolecular folding of the PCR product strand, resulting in secondary structures which subsequently slows down the hybridization kinetics. For instance, while PCRb (226 nt) works well in generating strong hybridization signals, PCRa (694 nt) does not produce sufficient intensities for differentiation.

4. Three-nucleotide repeat: Three-nucleotide repeats should be avoided, especially when they involve the SNP site as they can cause misalignment or 'slipping' of the PCR strands and probes. For instance, while N1Q has three-nucleotide repeats, N3Q does not. Therefore, N3Q is a better probe than N1Q in differentiation of *P. ginseng* and *P. quinquefolius*.

Chapter 4 On-Chip Nucleic Acid Amplification Test using Unamplified Genomic DNA

The information provided in the Introduction and Material and Methods herein are in addition to what has been presented in Chapter 1 and Chapter 2, respectively.

a) Introduction to SNP Detection Technology

LIDA (lesion-induced DNA amplification) is a method developed for SNP detection. There are various methods of differentiating between species with SNP sites. One method is to do a hybridization experiment where a DNA probe is hybridized to a PCR strand target containing the SNP site. There are various ways of determining whether the probe hybridization contains a perfect complementary match or a mismatch, namely, elevated temperatures, low ionic strength, wash by gold nanoparticles.[18] The hybridization method, however, requires a lot of optimization for a specific system.

SNP detection also allows us to differentiate between various species. DNA sequencing has become a popular alternative for species differentiation; while sequencing does not require extensive optimization, it does require samples of high purity in order to work.[115]

Another method capable of SNP detection is developed and it is called gold nanoparticle-based universal oligonucleotide microarray assay (GUMA). This method works by first producing a PCR product of the SNP-containing region like what would be performed in a standard probe-target hybridization. However, unlike the probe-target method in which a probe binds to a strand of this PCR product, it is used as a template in GUMA for a ligation reaction by *Taq* ligase. In this reaction, two probes hybridize to the SNP-containing region and if the PCR template and probes are perfectly complementary, then the ligase will ligate the two probes together. These probes also have universal primer hybridizing regions which allow for a second round of PCR amplification using labeled primers.[116]

Another method that has been developed is called arrayed primer extension (APEX). In this method, a short PCR product is made where the primers bind immediately upstream of the SNP site. The genotype of the SNP site is then identified by introducing ddNTPs, each labeled with a different fluorophore.[41] Such a test that combines nucleic acid amplification and detection is called nucleic acid amplification test (NAAT).

A major issue of these methods is that they all require external amplification by a technique, such as PCR, which also requires a thermocycler to perform. Given the high equipment cost and the tedious process, the applicability of PCR in commercial settings becomes limited. Therefore, isothermal amplification/detection techniques, which only need a hot water bath at a definite temperature, have become of interest.

There are many isothermal techniques for NAAT, such as loop-mediated isothermal amplification (LAMP)[117], rolling circle amplification (RCA)[118], strand displacement amplification (SDA)[119] and helicase-dependent amplification (HDA). HDA uses a

helicase, instead of a high temperature, to denature the templates/amplicons.[106,120] However, a subsequent hybridization step is still needed for differentiation.

An alternative to performing amplifications based on polymerases is to perform a ligation-chain reaction using ligases.[107] This method is called lesion-induced DNA amplification (LIDA); this is a two-cycle chain reaction which can do simultaneous amplification and differentiation isothermally.[108] Unlike polymerase, ligase is very sensitive to the proper base pairing at the ligation site which can be used as a tool for SNP differentiation. By introducing an abasic site to the ligation site, the ligated product will spontaneously denature in Cycle 1, and then the detection fragment is applied in Cycle 2, see Figure 4.1. We adapted the solution-phase LIDA to become the solid-phase LIDA and introduced the detection fragment in Cycle 1 which generated the fluorescence signal right away. Therefore, such a method eliminates the need for external amplification (e.g. by PCR or HDA).

In this chapter, solid-phase LIDA is developed and is used to identify the N2 site of SNP present in the DS gene of ginseng which differs between *P. ginseng* and *P. quinquefolius*.[97,98] This method has been applied to our microfluidic biochip (containing 16 parallel channels[42]) which is used to conduct amplification and differentiation with two advantages. First, this method allows for small volumes of the LIDA sample mixtures to be used. Second, the ligated product of the reaction will be bound to a solid support, which eliminates the requirement of markers conjugated on both strands of the ligation reaction needed for fluorescence resonance energy transfer (FRET) as employed for detection in the previous solution-phase method.[121] Furthermore, plant genomic DNA extracts obtained from ginseng samples were directly used as DNA templates for LIDA that was conducted at room temperature, which eliminated the need for even a hot water bath in this process.

b) Materials and Methods

All ss oligonucleotides and gBlock ds oligonucleotides used were purchased from IDT (Coralville, IA, USA). A plant DNA extraction kit (Qiagen) was used to extract the genomic DNA of ginseng samples. We had six ginseng powdered samples (X1-6) and seven ginseng root samples (2 Chinese, 3 American, 2 Korean) and one granular sample (Korean). T4 DNA ligase was obtained from NEB and Thermo Fisher (TF).

To decide on which SNP site to use, DINAMelt[29] was used to predict the Tm value of each of the DNA fragments used in the 2016 report (both perfectly complementary and 1 bp mismatch).[121] Next, similar length of DNA fragments corresponding to the N1 and N2 SNP sites in the DS gene were designed on DINAMelt,[29] see

Table 10 for more details. The fragments corresponding to the N2 site had the most similar Tm values of the ligated product (green-coloured entry #2 in

Table 10c) to those used previously (entry #2 in

Table 10a).[121] Even though the Tm values of three of the remaining four green coloured entries are higher for N2 than for the U of A sequences,[121] these N2 sequences are associated with the hybridization of the short fragments meaning that a higher Tm value is a good feature. On the other hand, the differences in Tm values between pm and mm are higher in N2 than in N1, indicating the higher differentiation capability of the former SNP site. Therefore, N2 was the locus chosen for the solid-phase LIDA reaction.

Table 10*:* Tm calculations used to decide on N2 instead of N1 for designing LIDA oligonucleotides. The table contains a list of all the duplexes that can form among the various DNA fragments and templates for the sequences used by Kausar *et al* at U of A (a),[107,121] by the proposed ginseng N1 site (b) and by the proposed ginseng N2 site (c). "Temp" referred to the *P. ginseng* sense template that helped produce the ligated product of "CompSNP Frag pm" and "CompAbFrag"; "AbCompTemp" refers to this ligated product. "SNP Frag mm" and "CompSNP Frag mm" are similar sequences to "SNP Frag pm" and "CompSNP Frag pm", respectively containing a single mismatching nt each, and "CompTemp" is a perfectly complementary sequence to "Temp". "Ab" in any of the names indicates that the strand contains an abasic position. When the N2 site is used in case c, Temp and CompTemp are the 17 nt sense strand around N2 of *P. ginseng* and *P. quinquefolius*, respectively; CompSNP Frag pm is Gin-NH_2 and CompSNP Frag mm is Quin-NH_2; CompAbFrag is Det; SNP Frag mm is QSF and SNP Frag pm is GSF. Finally, Frag is the same in either the N1 or N2 site (see Table 11). The Tm values highlighted in green pertain to the most relevant duplexes in the amplification reaction. Since DINAMelt interprets the abasic positions as a conventional mismatching bp, the Tms of the "Temp-AbCompTemp" duplexes are overestimated[122] for all three loci.

a) UofA Sequences	10mM Na, Mg, 37°C			
	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/molK)	Tm (°C)
Temp-CompTemp	-13.2	-129.6	-375.3	50.1
Temp-AbCompTemp	-12	-115.2	-332.8	48.3
Temp-Comp Frag	-6.4	-65.9	-191.9	29.8
Temp-CompAb Frag	-5.7	-59	-171.8	25.7
Temp-CompSNP Frag pm	-5.1	-55.2	-161.7	21.6
Temp-CompSNP Frag mm	-3.8	-44.9	-132.5	10.8
ComTemp-Frag	-5.3	-58.3	-170.9	23.5
ComTemp-SNP Frag pm	-5.8	-60.2	-175.4	26.3
AbComTemp-SNP Frag pm	-5.1	-55.2	-161.7	21.6
ComTemp- SNP Frag mm	-4.2	-47.8	-140.5	14.7
AbComTemp-SNP Frag mm	-4.2	-47.8	-140.5	14.7
b) Prop Ginseng Sequences (N1)	10mM Na, Mg, 37°C			
	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/molK)	Tm (°C)
Temp-CompTemp	-14.4	-128.8	-348.8	53.4
Temp-AbCompTemp	-12.7	-113	-323.5	50.5
Temp-Comp Frag	-6.2	-62.6	-181.7	28.8
Temp-CompAb Frag	-5.4	-54.7	-159.1	23
Temp-CompSNP Frag pm	-5.5	-51	-146.6	22.9
Temp-CompSNP Frag mm	-4.9	-43	-123	16.2
ComTemp-Frag	-5.1	-57.6	-169.2	22.6
ComTemp-SNP Frag pm	-6.4	-59.2	170.2	29.2
AbComTemp-SNP Frag pm	-5.5	-51	-146.6	22.9
ComTemp-SNP Frag mm	-4.9	-43	-123	16.2
AbComTemp-SNP Frag mm	-4.9	-43	-123	16.2
c) Prop Ginseng Sequences (N2)	10mM Na, Mg, 37°C			
	ΔG (kcal/mol)	ΔH (kcal/mol)	ΔS (cal/molK)	Tm (°C)
Temp-CompTemp	-14.5	-127.4	-364.1	53.7
Temp-AbCompTemp	-12	-110.8	-318.6	48.7
Temp-Comp Frag	-6.2	-50.3	-142.3	26.4
Temp-CompAb Frag	-5.1	-52.9	-154.2	20.9
Temp-CompSNP Frag pm	-5.8	-66.1	-194.5	27.1
Temp-CompSNP Frag mm	-4.1	-52.6	-156.4	15.8
ComTemp-Frag	-5.7	-54.8	-158.4	24.6
ComTemp-SNP Frag pm	-6.6	-68.2	-198.5	31.1
AbComTemp-SNP Frag pm	-5.8	-66.1	-194.5	27.1
ComTemp-SNP Frag mm	-4.1	-52.3	-155.5	15.6
AbComTemp-SNP Frag mm	-4.1	-52.3	-155.5	15.6

Immobilization: A 16-channel PDMS slab was first sealed onto an aldehyde-functionalized glass slide (75 mm × 50 mm) to form a microfluidic biochip. The microfluidic channels were filled with solutions containing Gin-NH_2 and Quin-NH_2 (25 μM) in the immobilization buffer (1.5 M NaCl, 0.15 M $NaHCO_3$) using suction, and the solutions in channels were incubated at room temperature for 1 h. Afterwards, the solutions were removed by suction, and the channels that were printed with Gin-NH_2 or Quin-NH_2 lanes were washed with the blank immobilization buffer. The PDMS slab was then removed, and the glass slide with the printed lanes was submerged in a $NaBH_4$ solution (~2.5 mg/mL $NaBH_4$, 0.75× PBS, 23.75% ethanol) for reduction for 15 min, washed with 1× PBS, and dried with nitrogen, as previously described.[42] The PDMS slab was washed, dried, and sealed back on the glass slide for a second time, but with the channels perpendicular to the printed lanes achieved in the first step.

Table 11: List of the oligonucleotides (ss and ds) used for the LIDA reaction. The sense sequence is shown in Figure 4.1 as the template in Cycle 1. The oligonucleotides in this table represents the following strands in

Table 10c: Detection Frag (CompAb Frag), Common Fragment (Frag), Gin-NH_2 (ComSNP Frag pm), Quin-NH_2 (ComSNP Frag mm), Gin Specific Fragment (SNP Frag pm), Quin Specific Fragment (SNP Frag mm), ligated Gin Specific Fragment and Common Frag (Temp), and ligated Gin-NH_2 and Detection Frag (AbComTemp).

DNA Fragment Name	**Sequence (5' to 3')**	**Functionalization**
Gin-NH_2 (antisense, ss)	CAATTTAAG	5' Amine
Quin-NH_2 (antisense, ss)	CAATTTAAA	5' Amine
InStd-NH_2 (antisense, ss)	CAATTTAAI	5' Amine
Detection Frag (antisense, ss, Det)	DCACTTTC	5' Phosphate, 5' Abasic#, 3' Biotin
Common Frag (sense, ss, Frag)	GAAAGTG	None
Gin Specific Frag (sense, ss, GSF)	TCTTAAATTG	5' Phosphate
Quin Specific Frag (sense, ss, QSF)	TTTTAAATTG	5' Phosphate
N2G (sense, ss)	GAATTTGAAA GTGTCTTAAA TTGATTTTCA A	5' Amine (Gin oligo template)
N2Q (sense, ss)	GAATTTGAAA GTGTTTTAAA TTGATTTTCA A	5' Amine (Quin oligo template)
gGin (Gin gBlock, ds)	CCATCAAGTA ATACCAAATT AATACTCCAT TCGTTCCATA ATATTTGATG ATTATTGAAT TTTTAATGTC AACTTAAGAG ATTTTTAATT TCAAAATTAA CATAGATTTT CTTTAAATAT ATACAGTGAT AATTAAATAT TGTAACTATC TAAAAAAAAA GTATTTTTCA TCTAAATTTT GAATTTGAAA GTGTCTTAAA TTGATTTTCA AAAGTCATAT AATTTGGAAC GGAGGGAGTA ACAAACATGA TTGAACATTG AGGTTAAAAA AAATAAATAA AATTTTGTGC TGTCCCTGCA TGCTAATTAA TTAAAAGGTG TTAACTGTTA TGTATGTAGA ATGAGGATGG TGGATGGGGA TCCTATATTG AGGGGCACAG CACGATGAT	Gin Standard Reference Template
gQuin (Quin gBlock, ds)	CCATCAAGTA ATACCAAATT AATACTCCAT TCGTTCCATA ATATTTGATC GTTATTGAAT TTTTAATGTC AACTTAAGAG ATTTTTAATT TCAAAATTAA CATAGATTTT CTTTAAATAT ATACAGTGAT AATTAAATAT TGTAACTATC TAAAAAAAAA GTATTTCTCA TCTAAATTTT GAATTTGAAA GTGTTTTAAA TTGATTTTCA AAAGTCATAT AATTTAGAAC GGAGGGAGTA ACAAACATGA TTGAACATTG AGGTTAAAAA AAATAAATAA AATTTTGTGC TGTCCCTGCA TGCTAATTAA TTAAAAGGTG TTAACTGTTA TGTATGTAGA ATGAGGATGG TGGATGGGGA TCCTATATTG AGGGGCACAG CACGATGAT	Quin Standard Reference Template

The abasic (D) moiety consists of a 1', 2' dideoxyribose.
"I" represents the inosine base which is a purine nucleoside.[123]

Ligation: A solution consisting of 2× ligation buffer, Detection Fragment (Det), Common Fragment (Frag), and either Gin or Quin Specific Fragment (GSF, QSF) (50 µM each) was made. The DNA template (gBlock ds oligonucleotides or genomic DNA) was first heated to 95°C for 5 min and then cooled on ice for 3 min; this was briefly centrifuged. The template (0.5 µL) was then mixed and combined with the mixture of

oligonucleotides (1.0 µL, 50 µM) and T4 DNA ligase (0.5 µL, 400 CEU/µL for NEB ligase, 5 U/µL for TF ligase) to give a final volume of 2 µL. This mixture was mixed thoroughly by pipetting up and down several times before being applied to the reservoirs for filling the channels of the chip by suction. The chip was incubated at rt for 2 hrs or 4 hrs, and then the solutions from the channels were removed by suction. The channels were then washed with 1.0% SDS and then 1× PBS before being filled with SA-Cy5 (50 µg/mL in 1× PBS) to be incubated for 15 min, followed by a 1× PBS/0.15% Tween-20 wash. The PDMS slab was then removed, and the glass slide washed with 1× PBS and dried with nitrogen.

Detection: The chip was scanned using a fluorescence imager (BioRad ChemiDoc); this instrument has similar resolution and sensitivity to the Typhoon imager used previously.[21,111]

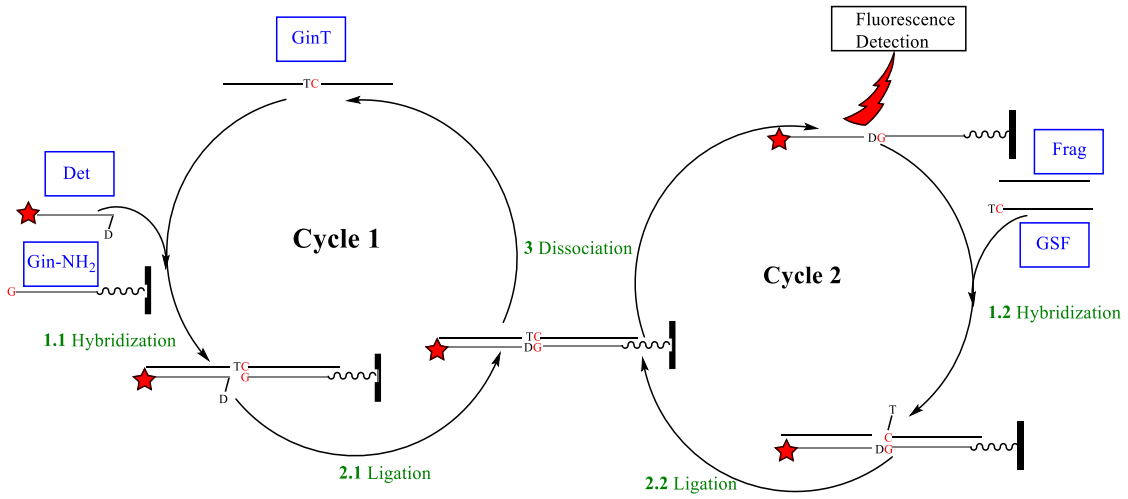


Figure 4.1: LIDA with two cycles in the ligation chain reaction. In Cycle 1, one template (A, GinT, sense) hybridizes (step 1.1) to the immobilized site-specific probe (Gin-NH_2, antisense) and causes ligation (step 2.1) of the detection fragment (Det, antisense). The duplex subsequently dissociates (step 3) to regenerate the template for Cycle 1 to repeat. Then, in Cycle 2, the immobilized long strand hybridizes (step 1.2) to site-specific fragment (GSF, sense) and causes its ligation (step 2.2) with the common fragment (Frag, sense). The duplex subsequently dissociates (step 3) to give an additional template (B, which is shorter than A) to enter Cycle 1. Therefore, Cycle 1 produces immobilized Det (fluorescence detection) to as many as the limiting amount of Det in the solution; Cycle 2 adds more templates (for Cycle 1) to as many as the amount of GSF or Frag, whichever is less.

c) Optimization Using ds gBlock Templates

As shown in Figure 4.1, in LIDA Cycle 1, one template (A, GinT, sense) hybridizes (step 1.1) to the immobilized site-specific probe (Gin-NH_2, antisense) and causes ligation (step 2.1) of the detection fragment (Det, antisense) to generate fluorescence signals. Then, in Cycle 2, the immobilized long strand hybridizes (step 1.2) to the site-specific fragment (GSF, sense) and causes its ligation (step 2.2) with the common fragment (Frag, sense); the duplex subsequently dissociates (step 3) to give an additional template (B, which is shorter than A) to enter Cycle 1. Strong fluorescence signals can be obtained by template GinT at the Gin-NH_2 lane in the presence of GSF, and by template QuinT at the Quin-NH_2 lane in the presence of QSF.

It was found that genomic templates, ss oligonucleotides as well as ds gBlock templates, were detected and differentiated by LIDA with both Cycles 1 and 2 (see data below). If only Cycle 1 was conducted (i.e. no GSF and Frag added), neither the genomic nor gBlocks (1 pg/µL) used as templates worked. The only template that did work with only Cycle 1 was ss oligonucleotides (100 nM but not 1 nM) (data shown in Appendix D Figure 7.27).

Several experimental conditions for LIDA have been optimized.

i) Reaction Time of LIDA (Need a 2,4,6 h Experiment)

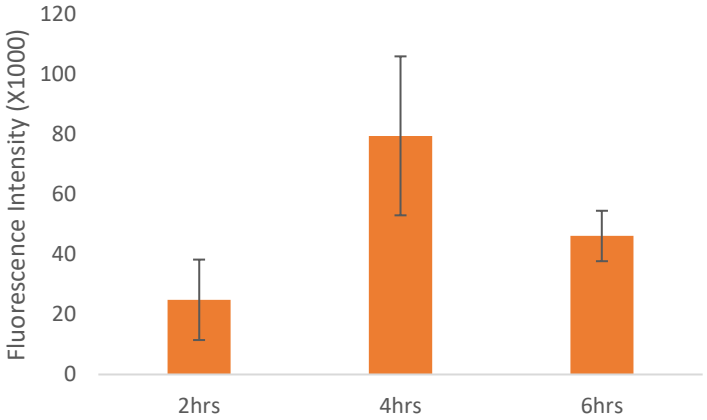


Figure 4.2: Effects of reaction time on LIDA. Three reaction times were used for the ligation reaction with Quin-NH_2 and Quin gBlock ds samples as templates using QSF. All the data bars are the averages of quadruplicates, and the error bars are the standard deviations. NEB ligase was used. Raw image data can be found in Appendix C (Figure 4.2).

The first aspect that was optimized was the LIDA reaction time where the three durations of 2, 4 and 6 hrs were tested using the gBlock ds Quin templates and Quin fragments. As it seen in Figure 4.2, there is a significant increase in fluorescence intensity between 2 h and 4 h, however, longer reaction time of 6 h did not help. In order to maximize LIDA intensities, 4 h was chosen to be the reaction time for subsequent LIDA reactions.

ii) Free NH_2-Oligo Added to the Ligation Mixture

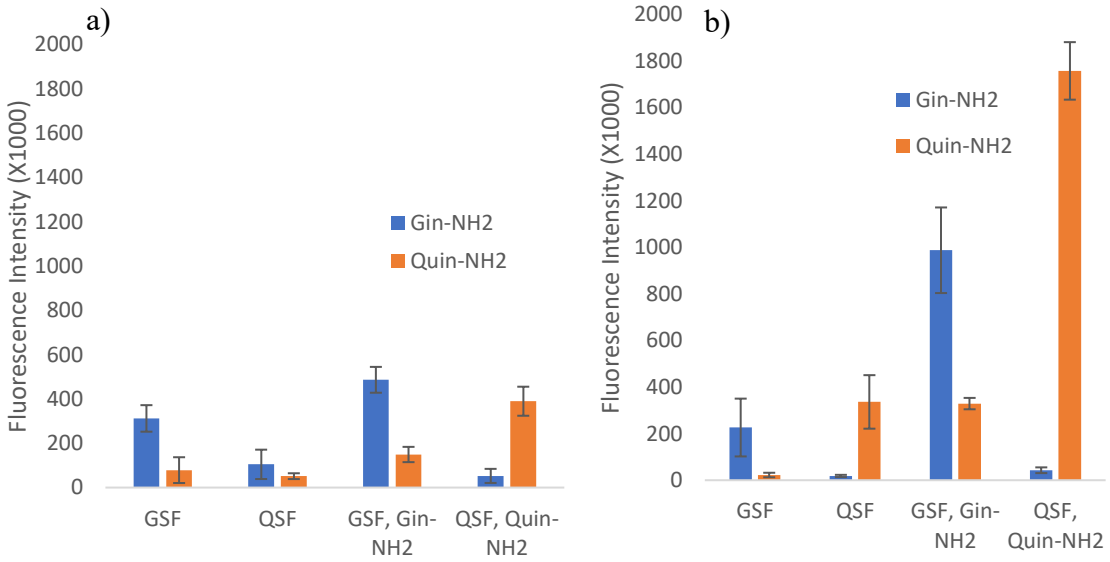


Figure 4.3: The effect of adding free-solution Gin-NH_2 or Quin-NH_2 in the ligation mixture. Gin-NH_2 or Quin-NH_2 (10 nM) were added to the ligation mixture, to be compared with controls containing no free-solution Gin-NH_2 or Quin-NH_2. The mixture was incubated for (a) 2 h, (b) 4 h. AmG2 genomic templates and 10 mM Mg^{2+} were used for all ligation reactions. Raw image data can be found in Appendix C (Figure 7.14).

The fluorescence intensities in initial LIDA experiments using the AmG2 template were relatively low. It was hypothesized that the low intensities were obtained because the template must hybridize to the immobilized oligonucleotide (Gin-NH_2/Quin-NH_2) in the very first cycle of the solid-phase LIDA. This hybridization that occurs in the solid phase is a slow process compared to the solution-phase reaction. Thus, if a small amount (10 nM) of Gin-NH_2/Quin-NH_2 was added to the LIDA solution mixture, the reaction will be jump started in the solution phase. The effect of adding this catalytic amount of -NH_2 oligonucleotide did help to increase the LIDA signals (Figure 4.3), and this effect was examined at 2 h (Figure 4.3a) and 4 h (Figure 4.3b). There is notably higher signal on the Quin-NH_2 lanes when QSF and Quin template are incubated for 4 h compared to 2 h. Similarly, there is high signal on the Gin-NH_2 lanes when GSF and Gin template are incubated for 4 h, but not 2 h. Thus, the immobilized Quin-NH_2 can identify the presence of the Quin sample.

Next, ds gBlock samples (gGin and gQuin) were used as templates in LIDA. As shown in Figure 4.4, the fluorescence intensity is the strongest at the Gin-NH_2 lanes for gGin when both the solution-phase oligonucleotide and immobilized oligonucleotide are Gin (GSF and Gin-NH_2), leading to a differentiation ratio (DR) of gGin from gQuin of 3.5 The value is calculated from the signals on the Gin-NH_2 lane (with GSF) by dividing that of gGin by that of gQuin. Likewise, although the fluorescence intensity is moderate at the Quin-NH_2 lanes for gQuin when both the solution-phase oligonucleotide (QSF) and immobilized oligonucleotide are Quin (Quin-NH_2), the differentiation ratio of 9.7 is reasonable (by dividing that of gQuin by that of gGin).

These differentiation ratios of 9.7 and 3.5 are good, though better ratios (i.e. 160 and 22, respectively) are obtained in LIDA when Cycle 2 was blocked and when ss oligonucleotide templates (100 nM) were used, see Appendix D Figure 7.27. This suggests that the inclusion of Cycle 2, though needed to increase signals, possibly reduces differentiation. Therefore, Cycle 1 produces immobilized Det up to as many as the limiting amount of Det in the solution, with differentiation dictated by Gin-NH_2. However, though Cycle 2 adds more immobilized Det up to as many as the amount of GSF or Frag, mismatch pairing might reduce differentiation.

This can be explained by the fact that when you only have Cycle 1, each iteration of the cycle is hindered if a mismatch is present since it is the same three strands that must hybridize together. When both cycles are utilized, even if you have a mismatch in Cycle 1 if a small amount of product manages to form from that cycle it will be perfectly complementary to the 2 fragments in Cycle 2. This will subsequently form a product that will be perfectly complementary to the Cycle 1 fragments thus resulting in the reaction losing its differentiation power. This makes optimizing the reaction time important so that it will form immobilized product only if the template and fragments are complementary.

In the scenario with GSF and gQuin, there is undesirable signal occurring at the Gin-NH_2 lanes (also seen in Figure 4.3); the presence of such a signal could be mistaken as the presence of any Gin template. When analyzing the complex that forms in Cycle 1 between the three strands (gQuin, Gin-NH_2 and Det), the SNP site forms a G-T mismatch base pair which is partially stabilizing allowing undesirable ligation to occur.

Another observation was that in situations where the two SNP fragments are not complementary (i.e. immobilize Gin-NH_2 with QSF and immobilized Quin-NH_2 with GSF), the fluorescence intensity is near background regardless of what template is present. This is explained by the fact that when you have mismatched SNP fragments, every iteration of the LIDA Cycle 1 and 2 must proceed through a mismatch. When both SNP fragments are complementary, then you only need a small amount of product to form from the first few cycles in order to cause the reaction to form more product through perfectly complementary base pairing.

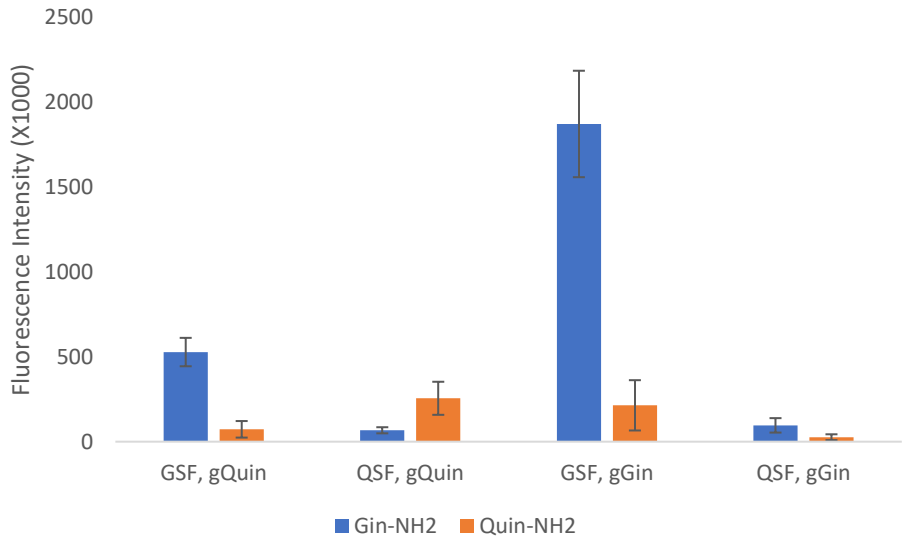


Figure 4.4: LIDA of gBlock templates. 100 CEU/μL NEB ligase was used and incubation time was 4 h. 10 nM Gin-NH_2 or Quin-NH_2, and 25 μM of GSF/QSF added. Raw image data can be found in Appendix C (Figure 7.15).

iii) Effect of Different Amounts of [Mg^{2+}]

When LIDA reactions were conducted using gBlock rather than genomic templates, it was seen that the QSF-gQuin intensity was significantly weak compared to the GSF-gGin intensities. The difference between these two LIDA reactions is that the reaction of QSF and gQuin involves a A-T bp at the SNP site while the reaction of GSF and gGin has a stronger G-C bp at the SNP site. This indicates that the association of the strands on the solid surface is the rate limiting step of the reaction. To help promote the association of the three oligonucleotides, the ionic strength was increased by adding more $MgCl_2$ (NaCl could not be used to increase solution ionic strength since Na^+ is an inhibitor of the ligase).[124]

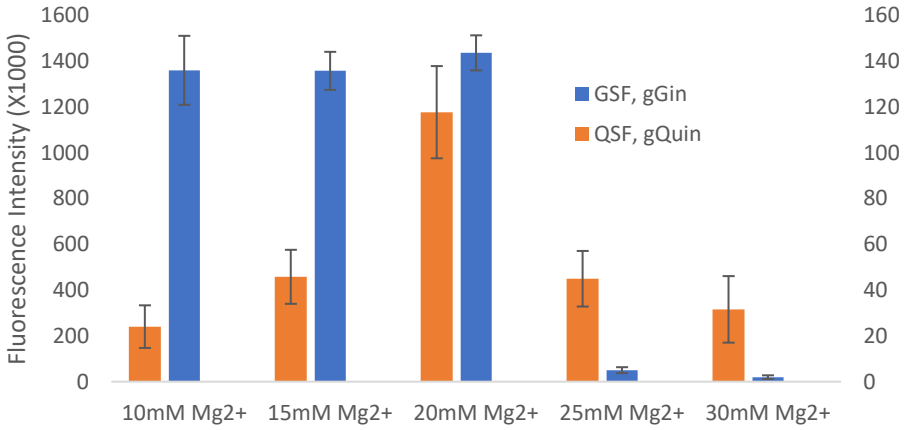


Figure 4.5: Effect of [Mg^{2+}] on LIDA. The ligation buffer already contains 10 mM Mg^{2+} and so additional Mg^{2+} was added to the buffer to result in the final concentrations of 15, 20, 25 and 30 mM. The blue bars are measured using the left vertical axis and the orange bars are measured using the right vertical axis. Raw image data can be found in Appendix C (Figure 7.16).

We observed that QSF, gQuin intensities increase with increasing [Mg^{2+}] up to 20 mM (Figure 4.5) after which the intensities decrease. Such a decrease in ligase activity at high magnesium ion content has been previously reported.[125] On the other hand, the GSF, gGin intensities are not affected by the [Mg^{2+}], from 10 mM to 20mM. Thus, 20mM Mg^{2+} was used for all subsequent LIDA experiments.

iv) Altering the Amount of QSF and Ligase

One persisting issue was that the signal from QSF on Quin-NH_2 lanes was always weak even when the template was perfectly complementary to the oligonucleotides. When the concentration of the QSF was increased from 25 µM to 30 µM, the intensity increased significantly as shown in Figure 4.6. Thus, subsequent LIDA experiments had the GSF and QSF fragment concentrations increased from 25 µM to 30 µM. The concentration of other fragments remained unchanged at 25 µM.

The effect of the concentration of ligase was also investigated (supporting data available in Appendix D Figure 7.23). It was determined that any dilution of the ligase results in a dramatic reduction in the fluorescence intensity. This is observed regardless of whether the diluent was DI water or 50% glycerol (v/v) (supporting data available in Appendix D Figure 7.24. As a result, the original concentration of 100 CEU/µL (final concentration) continued to be used.

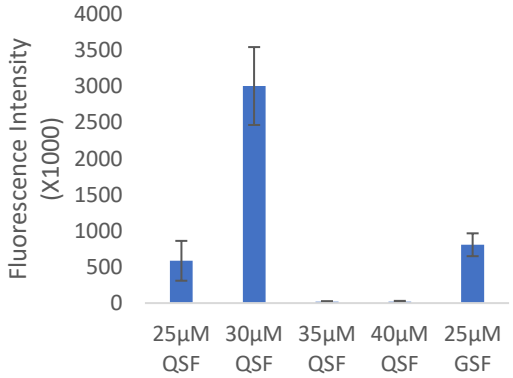


Figure 4.6: The effect of [QSF] on gQuin signals on Quin-NH_2 lanes. gQuin (1.0 pg/µL) was used as the template; Quin-NH_2 was added; 100 CEU/µL ligase was used with 4 h incubation. The last bar pertains to 25µM GSF used with 1.0 pg/µL gGin on Gin-NH_2 lanes. Raw image data can be found in Appendix C (Figure 7.17).

d) Analysis of 'Known' and Unknown Ginseng Samples

i) Pure ds Templates (gBlocks)

After optimizing the various factors governing the LIDA reaction, an experiment was performed to provide a reference of how a pure ds Gin sample, pure ds Quin sample and Gin/Quin mixed ds sample would appear. An issue was encountered with the T4 ligase first purchased from NEB, and we required to use the T4 ligase purchased from TF. The TF ligase was reported to be 2.5X more concentrated than the NEB counterpart. The TF ligase was tested by diluting it by 2 times using either DI water or 50% glycerol (v/v) and comparing the results obtained from these diluted ligases to those from the undiluted one. Only the undiluted ligase generated significant fluorescence intensity (supporting data

available in Appendix D Figure 7.25), thus undiluted TF ligase was also used for all gBlock and genomic ginseng sample analyses.

As seen in both Figure 4.7a and b, the fluorescence intensity is the highest when all three SNP-containing components are complementary to each other compared to when the template is a mismatch. Moreover, the mismatch intensity is higher for gQuin compared to gGin; both of these features have been observed previously (see Figure 4.4). The notable difference between the two ligase brands is that the NEB ligase did not allow the differentiation of the gGin and gQuin based on GSF (Figure 4.7a) while the TF ligase did (Figure 4.7b). Moreover, for the mixed gGin/gQuin templates, the NEB ligase produced a near background noise intensity while the TF ligase resulted in an equal intensity for the GSF and QSF bars comparable to the pure gGin or gQuin. The likely explanation for this is that since the TF ligase (5 U/µL) is of a higher concentration than the NEB ligase (400 CEU/ µL or 2 U/µL), the TF ligase can handle the lower concentrations of the complementary gBlock template in the mixtures (total gBlock concentration is the same as the pure samples but each subtype is half the concentration) better than the NEB ligase. The higher enzyme concentration in the TF ligase also results in higher overall fluorescence intensities as can be seen by examining the scale of the y-axis.

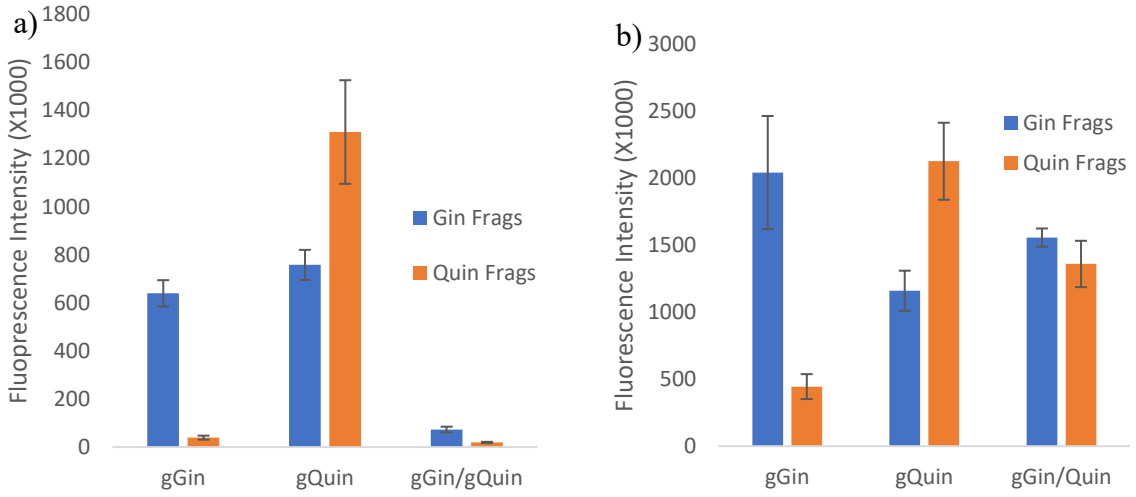


Figure 4.7: Intensity for 100% gGin, 100% gQuin, and a 50% mix of each (gGin/Quin). The LIDA mixture was composed of 1 µL 2× Oligo/buffer solution (Det, Frag (50 µM each), and 60 µM GSF or QSF, Gin-NH_2 or Quin-NH_2 (20 nM), 2× buffer, 20 mM $MgCl_2$), 0.5 µL template and 0.5 µL ligase (2 µL final volume). The ligation reaction was 4 h. a) 100 CEU/µL NEB ligase b) 1.25 U/µL TF ligase. Raw data can be found in Appendix C (Figure 7.18).

An examination of the list of ingredients of the ligases obtained from the two vendors (NEB and TF), as shown in Table 12, did not reveal substantial differences. So, I think it is the concentration of the enzyme, rather than the buffer ingredients, that matters because experiments employing a lower concentration of TF ligase (diluted from 1.25 to 0.625 U/µL) provided negative results that could not be distinguished from the background.

Table 12: The lists of buffer ingredients for T4 ligases obtained from the two vendors NEB and TF.

1× NEB T4 DNA Ligase Reaction Buffer (for 400 CEU/µL or 2 U/µL T4 DNA ligase).
50 mM Tris-HCl (pH 7.5 @ 25°C)
10 mM $MgCl_2$
1 mM ATP
10 mM DTT

1× TF T4 DNA Ligase Buffer (for 5 U/µL T4 DNA ligase)
40 mM Tris-HCl (pH 7.8 @ 25°C)
10 mM $MgCl_2$
0.5 mM ATP
10 mM DTT

To improve the differentiation of the gGin and gQuin based on GSF using TF ligase, it was seen that 2 h incubation has better differentiation than 4 h (data available in Appendix D Figure 7.26). The LIDA results obtained at 2 h are shown in Figure 4.8. These results lead to DR of 5.3 for Gin and 40.8 for Quin, which are better than the previous values of 3.5 and 9.7, respectively, found on Figure 4.4. Subsequent LIDA experiments that used gBlock templates and TF ligase were performed with a 2 h incubation.

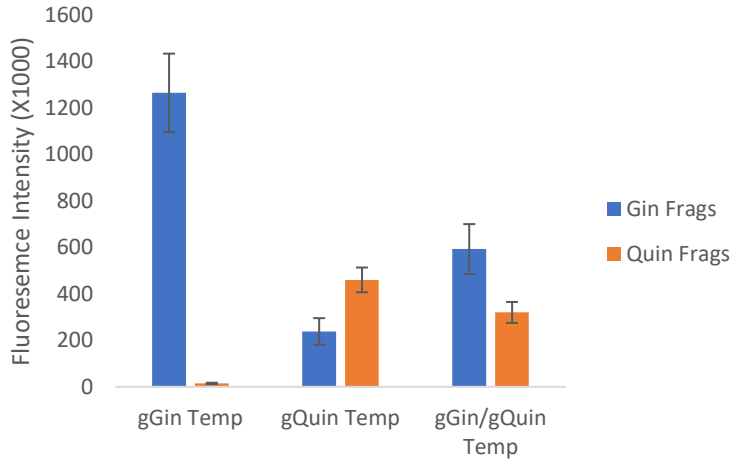


Figure 4.8: Intensity for 100% gGin, 100% gQuin, and a 50% mix of each (gGin/Quin). The LIDA mixture was composed of 1 µL 2× Oligo/buffer solution (Det, Frag (50 µM each), and 60 µM GSF or QSF, Gin-NH_2 or Quin-NH_2 (20 nM), 2× buffer, 20 mM $MgCl_2$), 0.5 µL template and 0.5 µL TF ligase (2 µL final volume) and was incubated for 2 h. Raw data can be found in Appendix C (Figure 7.19).

ii) Gin/Quin Ratios of Unknown Ginseng Samples

After optimization of the TF ligase, unknown genomic ginseng samples X1 to X6 were analyzed (Figure 4.9).

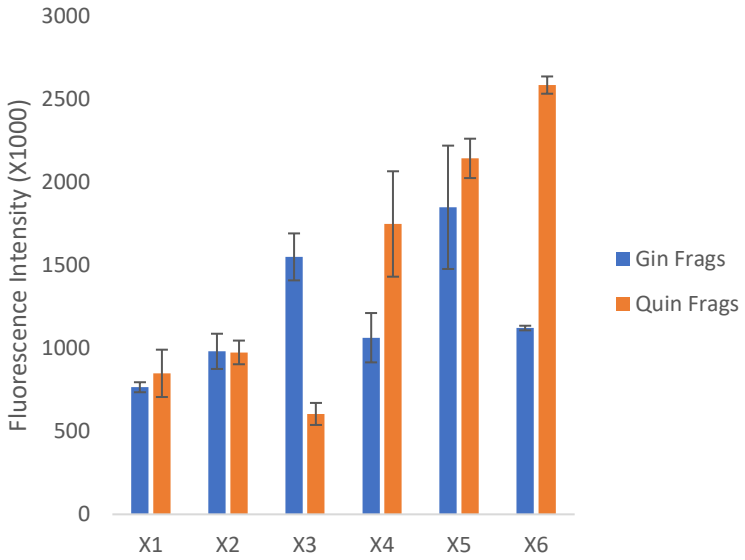


Figure 4.9: Differentiation of ginseng samples X1 to X6. Genomic templates used; 10 nM Gin-NH_2 or Quin-NH_2 added, 30 µM GSF or QSF, 25 µM Det, 25 µM Frag, 10 mM [Mg^{2+}] (added), 1.25 U/µL TF ligase, 2 h incubation. Raw data can be found in Appendix C (Figure 7.20).

In order to conduct species authentication, we create a parameter called Gin/Quin ratio. This is the ratio of fluorescence intensity 1 to intensity 2, where intensity 1 was obtained on Gin-NH_2 lanes when GSF was in the LIDA buffer and intensity 2 was obtained on Quin-NH_2 lanes when QSF was in the LIDA buffer. The Gin/Quin ratios are shown in Table 13.

Table 13: Differentiation Gin/Quin ratios calculated from Figure 4.9.

X1	**X2**	**X3**	**X4**	**X5**	**X6**
0.90	1.01	2.56	0.61	0.86	0.43

A high ratio (>2.0) in Table 13 (e.g. X3) indicates that the sample is *P. ginseng*-like. On the other hand, a low ratio (<0.5) (e.g. X4 and X6) indicates that the sample is *P. quinquefolius*-like. This would make the remaining three samples (0.5 < ratio < 2.0) appear to be hybrids of the two species.

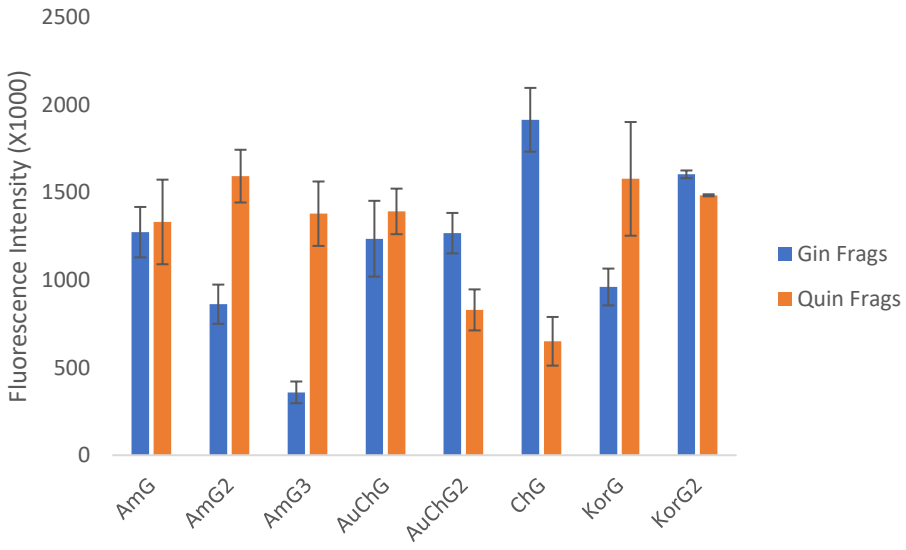


Figure 4.10: LIDA of the eight ginseng root samples with LIDA. Though AmG, AmG2 and AmG3 were claimed to be American ginseng, only AmG2 and AmG3 can live up to the claims. While AuChG, AuChG2, KorG, and KorG2 and were expected to be Chinese ginseng, only AuChG2 and ChG fit better to the claimed identities. The reaction was carried out using 1.25 U/μL TF ligase and 2 h incubation. Raw data can be found in Appendix C (Figure 7.20).

iii) NGS of Ginseng Samples

NGS was done using the Illumina platform by Genewiz. FASTQ files were acquired from Genewiz and analyzed using the Burrows-Wheeler Aligner (BWA).[126] BWA was used to align the R1 (sense) FASTQ sequences to the expected *P. ginseng* 126 nt sense strand (in FASTA format), generating the sequence alignment map (SAM) file. The newly formed SAM files were then visualized using Integrated Genomic Viewer (IGV)[127] and the *P. ginseng*/*P. quinquefolius* ratios were determined using the interface.

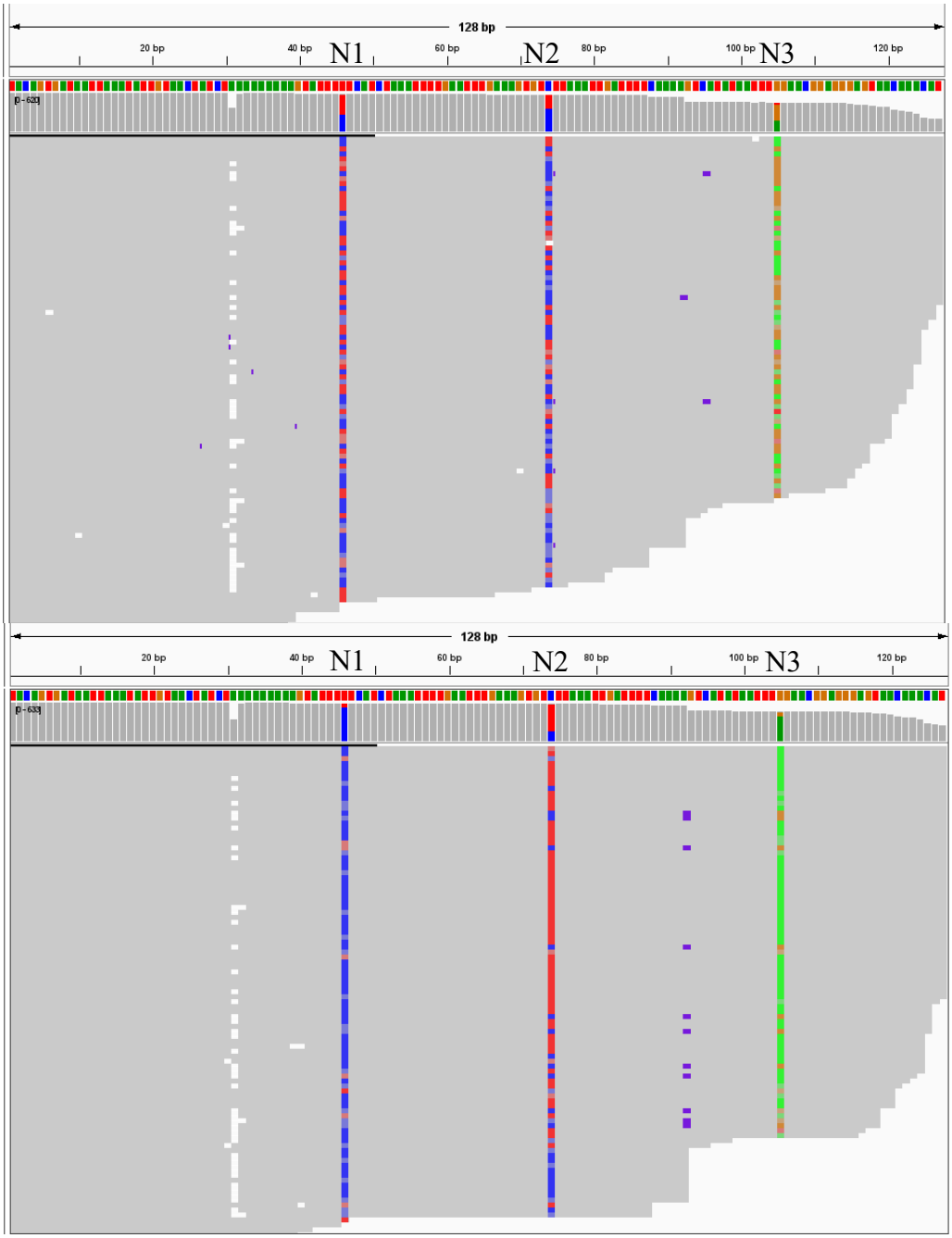


Figure 4.11: IGV image of the ChG (top) and KorG (bottom) sequences where the colours represent C (blue), T (red), G (orange), A (green), deletion (white) and insertion (purple). The row of coloured pixels at the top represents the *P. ginseng* reference sequence. The bars below the reference sequence represent the cumulation of all nucleotides at the specific position; any position that deviates by ≥20% relative to the reference will show up coloured in. Below the bars are the sequences themselves listed on order of decreasing length. Any position that was coloured in the bars is also coloured in within the sequences. The N1, N2 and N3 sites are labeled above the reference sequence. The ChG image contain 620 sequences and the KorG image contains 633 sequences.

The IGV images of the most *P. ginseng*-like sample (ChG) and the most *P. quinquefolius*-like sample (KorG) are presented in Figure 4.11. The bases at the three SNP sites of N1, N2 and N3 are depicted as three columns in the images. Looking at the top image of ChG, the sample appears to have more of the *P. ginseng* genotype: T, C, G

(red, blue, orange) than the *P. quinquefolius* genotype: C, T, A (blue, red, green), but the sample does appear to be heterozygous, nonetheless. What is also seen is that in front of the N1 SNP site there is a 1 nt deletion (shown in white) in many of the sequence variants, and in between the N2 and N3 SNP sites, there are the insertion (AACAAACAATA/GATT/C, where the underlined sections are either A or G and T or C, respectively) as shown in Appendix F, Table 23. On the other hand, looking at the bottom image in Figure 4.11 (KorG), the heterozygous nature of the samples is also visible though with more *P. quinquefolius* genotypes than *P. ginseng* genotypes. It is also worth noting that the same insertions are present in the KorG image as were seen in the ChG image between the N2 and N3 SNP sites. Furthermore, these insertions (shown in purple) are observed on sequences that have the SNP genotype of CCG which is a new genotype that has not been previously reported.[97] This newly discovered genotype also has the 1 nt deletion in front of the N1 SNP site. This CCG genotype containing the insertion (shown in purple) was also seen in the top image.

Table 14: The ratios (G/Q) of the relative abundance of the *P. ginseng* and *P. quinquefolius* genotypes at the N1, N2 and N3 SNP sites of the 14 ginseng samples obtained from NGS through analysis using BWA and IGV. The samples are classified as *P. ginseng* (G), *P. quinquefolius* (Q), and heterozygous (H).

	N1	N2	N3	
X1	0.31	0.42	0.26	Q
X2	0.25	0.38	0.19	Q
X3	0.63	1.06	0.82	H
X4	0.13	0.31	0.20	Q
X5	0.36	0.41	0.23	Q
X6	0.18	0.37	0.17	Q
AmG	0.26	0.58	0.34	Q
AmG2	0.55	0.71	0.58	Q
AmG3	0.63	0.84	0.65	H
AuChG	0.54	0.99	0.81	H
AuChG2	1.43	1.52	1.68	G
ChG	1.12	1.62	1.47	G
KorG	0.12	0.40	0.17	Q
KorG2	0.43	0.40	0.32	Q

The *P. ginseng*/*P. quinquefolius* genotype ratios were also calculated and compiled for all three SNP sites, see Table 14. By observing all 14 samples in Table 14, we determine the ratio is the highest at the N2 site for all the samples except AuChG2 and KorG2. Based

on these observations, the samples are classified as *P. ginseng* (G), *P. quinquefolius* (Q), and hybrid (H).

The discovery of the CCG SNP genotype introduced the question of how abundant it was in comparison to the *P. ginseng* (TCG) and *P. quinquefolius* (CTA) genotypes. A Java program (script available in Appendix F) was written which identified the three SNP sites by searching in the FASTQ files the conserved 4-8 nt sequence directly in front on the respective SNP site. The N2 SNP sites was used to verify that the Java program is a viable method for analyzing the FASTQ files. The abundances of the two genotypes determined using BWA was compared to the abundances using the Java program for the same samples (Figure 4.12). While there are some small differences in the abundances especially for the samples with a higher *P. ginseng* composition (e.g. AuChG, AuChG2, ChG), there is a general agreement between the results obtained from the two methods of FASTQ file analysis.

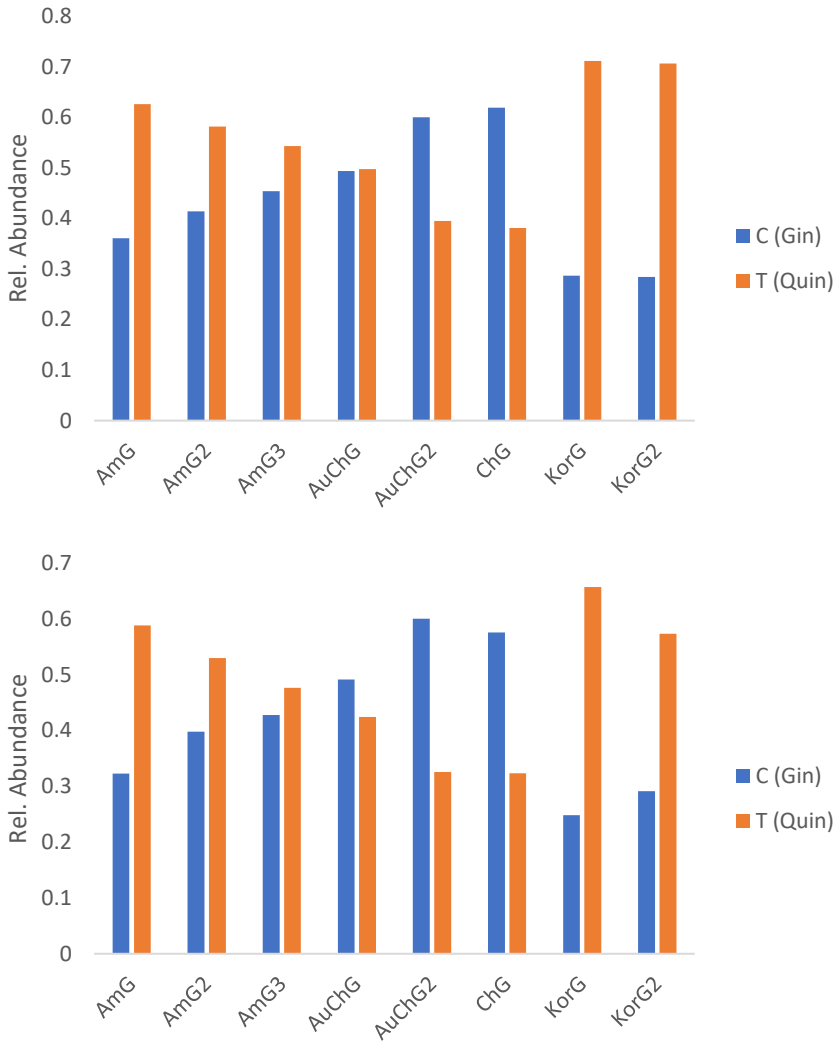


Figure 4.12: NGS results of the N2 SNP site for AmG, AmG2, AmG3, AuChG, AuChG2, ChG, KorG, and KorG2 using BWA/IGV (top) and the Java program (bottom) where *P. ginseng* contains a C base, and *P. quinquefolius* contains a T base. All intensities have been normalized to the total number of sequences for each sample. Raw data for these root samples and for X1-X6 (powder samples) can be found in Appendix F, Table 20, Figure 7.28.

The results from Figure 4.12 have shown that the Java program is a viable approach for use to quantify the different genotypes. The CCG genotype that was described for ChG and KorG in Figure 4.11 is also identified in other samples, and this genotype is actually more abundant than the *P. ginseng* genotype: TCG in many of the samples, see Figure 4.13. The only two samples that have a higher abundance of the *P. ginseng* genotype (compared to the *P. quinquefolius* genotype: CTA) are the AuChG2 and ChG. On the other hand, some samples (e.g., X6 and KorG) have very little *P. ginseng* genotype; the presence of the CCG genotype gives them an apparently higher *P. ginseng* character as shown in Figure 4.12 based on N2 alone. The CCG genotype accounts for a significant portion of all 14 samples which has made the N1, but not N2, site difficult to get conclusive results for many of the samples, especially using N1Q for NAT based on hybridization.

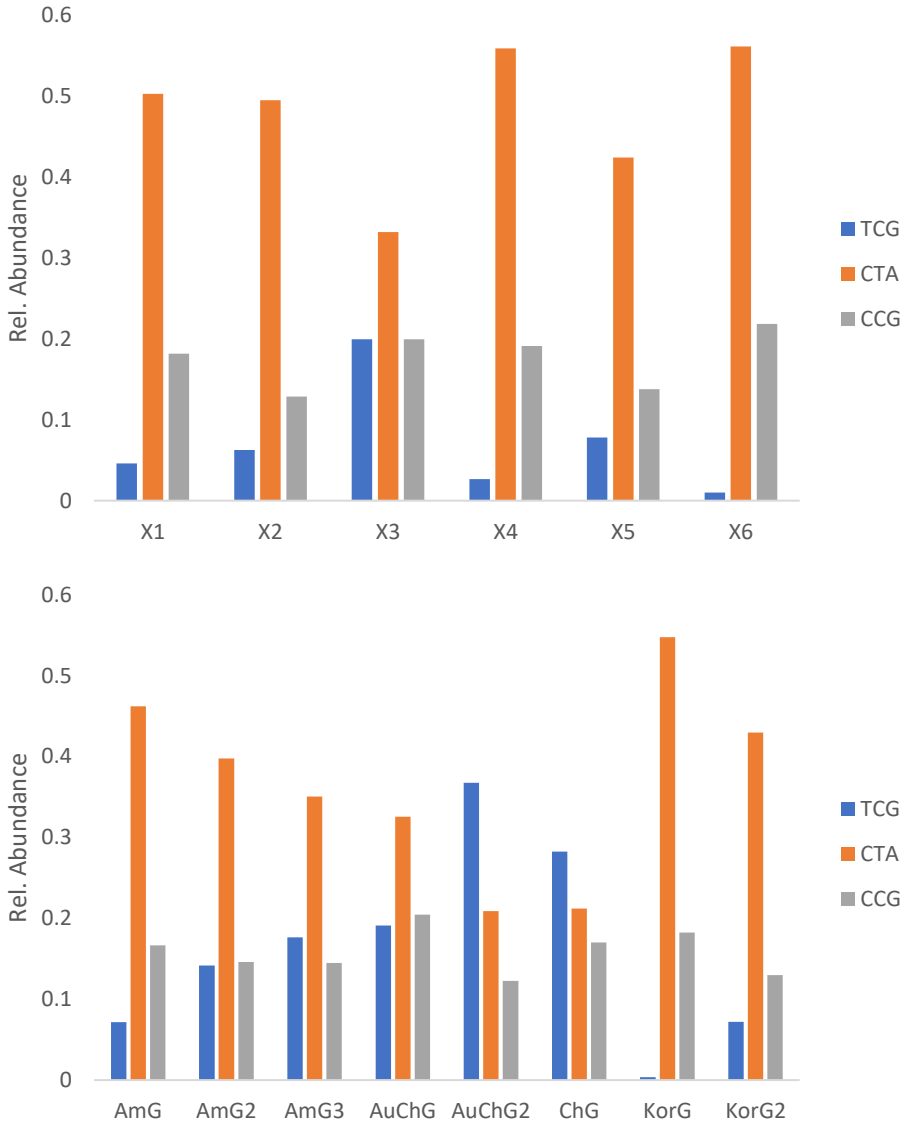


Figure 4.13: NGS results of the N1, N2 and N3 SNP sites of X1-X6 (top) and AmG, AmG2, AmG3, AuChG, AuChG2, ChG, KorG, and KorG2 (bottom). *P. ginseng* contains the TCG genotype and *P. quinquefolius* contains the CTA genotype. The new genotype is indicated as CCG. All intensities have been normalized to the total number of sequences for each sample. Raw data can be found in Appendix F (Table 22).

The discovery of the CCG genotype along with the deletions and insertions have shown that N1 is not a good choice for genotypic authentication of *P. ginseng* and *P. quinquefolius*. We have found many allelic and intraspecies variations that were not previously reported.[97] Future work on the genotypic authentication of ginseng should be performed on a locus that has low allelic/intraspecies variation.

e) Consideration for an Internal Standard

There was an interest to have an internal standard for LIDA in the channels where samples flowed in, and so a third immobilized oligonucleotide (InStd-NH_2) that would form a ligated product in addition to the two sample-specific ones (i.e. Gin-NH_2 and

Quin-NH_2) would be helpful. This internal standard should form the ligated product regardless of what combinations of SNP oligonucleotides (i.e. GSF and QSF) and templates (i.e. Gin or Quin) is used. Inosine (I) contains a naturally occurring hypoxanthine base that is resulted from the deamination of adenosine that is capable of wobble base pairing with any of the four bases (e.g. C, U or T).[123] It was hypothesized that having an immobilized strand containing an inosine at the SNP site would allow the LIDA reaction to proceed to a similar extent regardless of what combination of template and SNP oligonucleotide (GSF/QSF) is used in the solution-phase.

What was observed was that regardless of what template (i.e. gGin and gQuin) was used, the fluorescence intensity at the InStd-NH_2 lane was always high when GSF was used and always weak when QSF was used (Figure 4.14a). In order to explain this observation, the stability of the base pairing of inosine (I) was examined. In each cycle of LIDA, the reaction in the GSF channels proceeds through an I-C bp, while the reaction in the QSF channels proceeds through an I-T bp. Thus, the results indicate that the ligase can ligate Frag with GSF (TCTTAAATTG) when hybridized to InStd-NH_2: CAATTTAAI based on I-C pairing, but not with QSF (TTTTAAATTG) based on I-T pairing. These are compared with the pairings in the ligation of Frag with GSF when hybridized to Gin-NH_2: CAATTTAAG (i.e. by G-C pairing), and with QSF when hybridized to Quin-NH_2: CAATTTAAA (i.e. by A-T pairing). Therefore, we examine various enthalpies of pairing (E_p) of these pairings as shown in Figure 4.14b, Table 15, and found that the E_p value of an I-T bp is very close to that of an A-T bp,[128] and yet, the ligase cannot ligate Frag with QSF to a large extent when hybridized to InStd-NH_2 . This means that it is not thermodynamics alone that determines what bps can or cannot be paired for ligation.

To consider kinetics, we next examine the pairing of the bases. In Figure 4.14b top, what is seen is that in the two Watson-Crick bps (G-C and A-T), you have H-bonding on positions 1 and 6 of the A base with positions 3 and 4, respectively, of T, and you have H-bonding on positions 2, 1, 6 of G with positions 2, 3, 4, respectively, of C (for numberings, see Figure 4.14c. In Figure 4.14b bottom left, in the I-C wobble bp, you have H-bonding on positions 1 and 6 of inosine with positions 3 and 4, respectively, of C. However, in Figure 4.14b bottom right, the I-T wobble bp consists of the H-bonding formed between positions 1 and 6 of inosine and positions 2 and 3, respectively, of the pyrimidine (T or U), which results in the two bases being staggered. Since DNA ligase is known[129] to have low tolerance for conformation abnormalities at the ligation site, this staggered arrangement of the I-T bp may be the likely cause of the poor performance of inosine in the LIDA reaction when used with the QSF.

Thus, inosine did not function as a proper internal standard due to its incompatibility with QSF when the ligase was used. This means that LIDA does not only require the bases at the SNP site to base pair but also require them to base pair with a Watson-Crick style alignment for the ligase to work; the ligase does not tolerate a staggered alignment of the bases. Therefore, since cytosine (C) has an H-bond acceptor and H-bond donor at positions 3 and 4, respectively, while thymine (T) has H-bond donor and H-bond acceptor

at positions 3 and 4, respectively, it is not possible to have a universal purine base that will base pair with both C and T (U) in an arrangement where neither of the two paired bases is staggered.

Table 15: Enthalpies of pairing of Watson-Crick and wobble base pairs.[128]

Base Pair	**Enthalpy of pairing (E_p) (kcal/mol)**
G-C	-11.71
A-T	-5.86
I-C	-10.04
I-T	-6.03

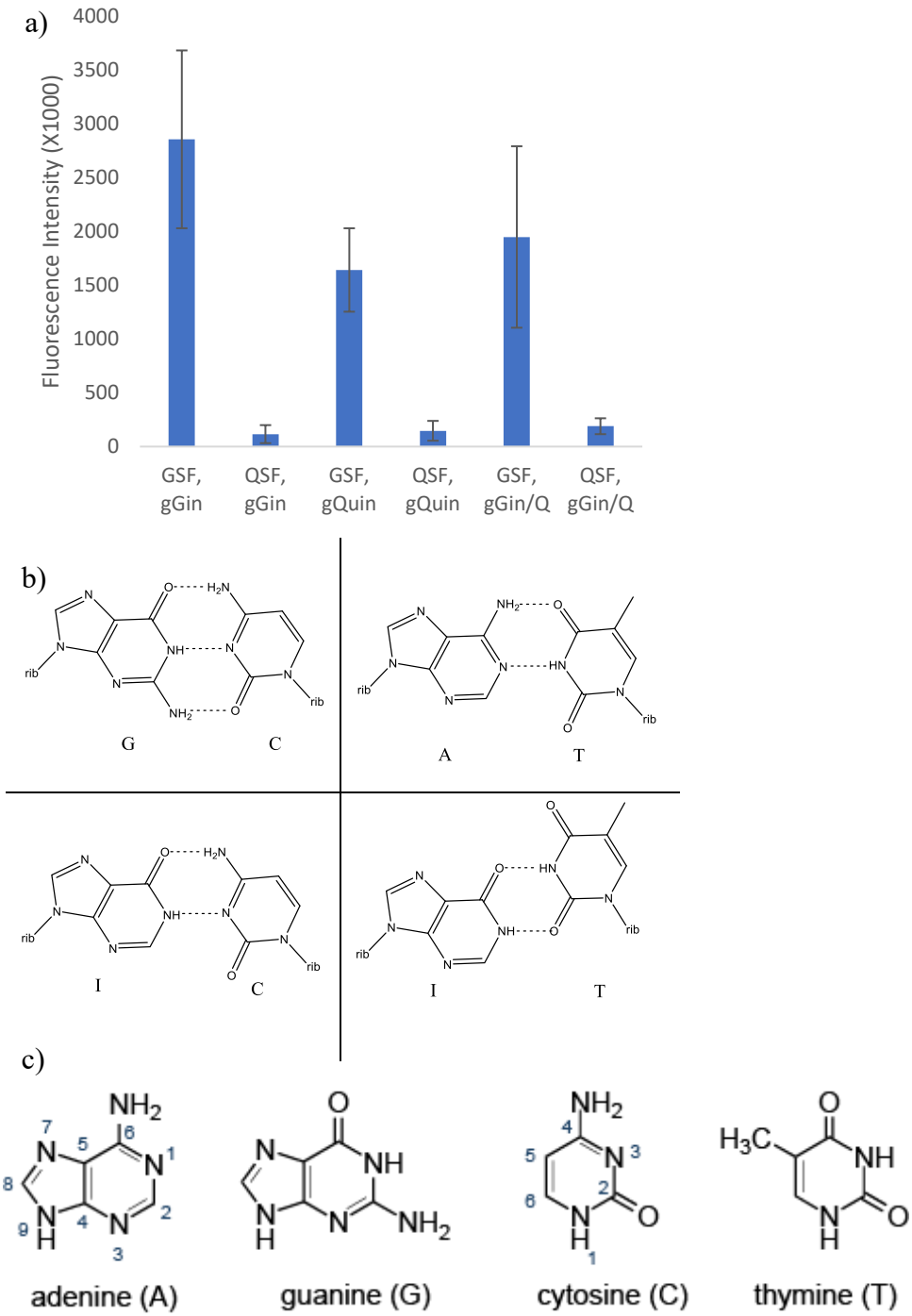


Figure 4.14: a) Fluorescence intensities at the immobilized fragment InStd-NH_2 used in combination with LIDA solutions containing 30 µM GSF or QSF, 25 µM Frag, 25 µM Det and 1.0 pg/µL gBlock templates. Raw image data can be found in Appendix C (Figure 7.21, Figure 4.14). b) Watson-Crick G-C and A-T base pairs (top) and wobble I-C and I-T base pairs (bottom). Ribose on the nucleotides is indicated by "rib". c) Numbering adopted for purines (A, G) and pyrimidines (C, T).

f) Conclusion

The results presented here have shown that the NAAT based on LIDA can be used in a microfluidic system as a means to differentiate between ginseng species based on the N2 SNP site present in the DS gene of the ginseng genome. This method allows unamplified genomic material to be used directly as the DNA template in an isothermal room temperature reaction, thus eliminating the need for external DNA amplification and a costly thermocycler. The methodology was used to identify six unknown ginseng powder samples, revealing that one is *P. ginseng*-like (X3), two are *P. quinquefolius*-like (X4, X6) and the remaining three are hybrids of the two species (X1, X2, X5). Further tests were conducted to examine various ginseng root and ginseng tea granular samples.

It was also observed that regardless of what DNA template is used (ss oligonucleotide, ds gBlock or genomic), the success of the LIDA reaction is more dependent on both, rather than just one, of the two complementary hybridizations, i.e. Cycle 1: Gin-NH_2 complementary to *P. ginseng* template, and Cycle 2: Gin-NH_2 complementary to GSF. In addition, a fragment containing inosine at the SNP site was attempted to be used as an internal standard, but this did not work in our system when T (or U) was involved. It was revealed in our study that this inosine fragment is only compatible with GSF, but not with QSF because the staggered I-T wobble bp in the latter case is not compatible with the T4 ligase.

Chapter 5 Conclusions and Future Work

In this thesis, three methods for nucleic acid test (NAT) have been examined: the probe-target method (Method 1), target-probe method (Method 2), and three-strand complex method (Method 3). The hybridization reactions are achieved by either the SF or dynamic condition. Each of the three methods were optimized using asymmetric PCR products of the DS gene of *P. ginseng* and *P. quinquefolius*. The method that gave the best differentiation was the three-strand method (Method 3) using the 3'-NH_2 P8' as the capture strand, the 226 nt PCR strands and the N2G and N3Q probes for detection. This Method 3 works the best because the 3'-NH_2 capture strand positions the PCR strand to face down towards the solid surface. This slows down the kinetics of the detection probe hybridization by making the PCR strand less accessible to the probes. This in turn favours the formation of the perfectly complementary over the mismatch duplexes.

Instead of extending the three-strand method to include an on-chip DNA amplification step to develop a nucleic acid amplification test (NAAT), we develop a totally new method for NAAT.

A new method for on-chip NAAT has been developed by using an isothermal DNA amplification method called LIDA in the solid phase. This method has been successfully employed to differentiate between *P. ginseng* and *P. quinquefolius* using unamplified genomic ginseng samples as a template. It was revealed that the success of the two cycles of the LIDA reaction is more dependent on both, rather than one of the two complementary hybridizations, i.e. *P. ginseng* template complementary to Gin-NH_2, and Gin-NH_2 complementary to GSF. Furthermore, through the use of the inosine nucleotide, it was determined that T4 ligase not only requires the SNP bases to be complementary to each other but also requires them to have the Watson-Crick, but not wobble, style of base pairing.

LIDA was used to assay all powder, root and granular ginseng samples. Their authentication by NAT and NAAT, which was confirmed by NGS data, was shown in Table 16. It is realized that pure *P. ginseng* is rare, though more samples are *P. quinquefolius*, and many samples are hybrids.

Table 16: Comparison of authentication results of powder and root ginseng samples.

	NAT	**LIDA**	**NGS**
X1	H	H	QH
X2	H	H	QH
X3	GH	GH	H
X4	Q	Q	Q
X5	H	H	QH
X6	Q	Q	Q
AmG	No Data	H	QH
AmG2	Q	Q	Q
AmG3	Q	Q	QH
AuChG	H	H	H
AuChG2	No Data	G	G
ChG	G	G	G
KorG	Q	Q	Q
KorG2	Q	Q	Q

G: *P. ginseng*; Q: *P. quinquefolius*; H: Hybrid; GH: *P. ginseng*-like; QH: *P. quinquefolius*-like

NGS results should be more reliable because all DNA sequences, up to 28 million sequences in a sample are analyzed and tabulated (see Appendix F).

A new genotype (CCG) has been uncovered in all 14 samples that were analyzed by NGS. In addition, there is a notable 1 nt deletion in front of the N1 SNP site and an insertion between the N2 and N3 SNP sites in many of the sequences. Fortunately, the only probe that is impacted by the insertion is the LN3Q which was examined but not used extensively in this thesis. However, these findings of the new genotype and deletion/insertion indicate the challenge that the previously reported *P. ginseng* and *P. quinquefolius* genotypes[97] did not uncover all the diversity that is present in the DS gene of ginseng. The challenge was partly resolved by using pure sequences such as ss oligonucleotides in NAT, or ds gblocks in LIDA. Future work on genotypic authentication of ginseng (or any other species) should be done only after verifying that the locus being selected for the project has low allelic variation. We believe the solid-phase LIDA developed in this thesis can be applied to detect any well-validated locus.

Future work on the on-chip LIDA method can include examining the use of a poly-T spacer on the immobilized strand to study the effect of the spacing between the LIDA-active sequence and the chip surface. In addition, the immobilized fragment density can

be optimized for the on-chip LIDA. The 25 μM concentration used was adapted from the NAT methodology but it may not necessarily be the optimal concentration for on-chip LIDA. Another aspect that can be examined is the stability of the NH_2 strands immobilized on the glass surface. Rather than performing the LIDA reaction immediately after the immobilization, the reaction can be performed several weeks after the immobilization to examine whether there is any deterioration of the immobilized strands.

Chapter 6 References

(1) Chiu, S. K.; Mandy, H. S. U.; Ku, W. C.; Tu, C. Y.; Tseng, Y. T.; Lau, W. K.; Yan, R. Y.; Ma, J. T.; Tzeng, C. M. Synergistic Effects of Epoxy- and Amine-Silanes on Microarray DNA Immobilization and Hybridization. *Biochem. J.* **2003**, *374* (3), 625–632.

(2) Woodruff, K.; Maerkl, S. J. A High-Throughput Microfluidic Platform for Mammalian Cell Transfection and Culturing. *Sci. Rep.* **2016**, *6*.

(3) Jung, D. H.; Kim, B. H.; Ko, Y. K.; Jung, M. S.; Jung, S.; Lee, S. Y.; Jung, H. T. Covalent Attachment and Hybridization of DNA Oligonucleotides on Patterned Single-Walled Carbon Nanotube Films. *Langmuir* **2004**, *20* (20), 8886–8891.

(4) Le Berre, V. Dendrimeric Coating of Glass Slides for Sensitive DNA Microarrays Analysis. *Nucleic Acids Res.* **2003**, *31* (16), 88e – 88.

(5) Fixe, F. Functionalization of Poly(Methyl Methacrylate) (PMMA) as a Substrate for DNA Microarrays. *Nucleic Acids Res.* **2004**, *32* (1), 9e – 9.

(6) Ranjbar, R.; Behzadi, P.; Najafi, A.; Roudi, R. DNA Microarray for Rapid Detection and Identification of Food and Water Borne Bacteria: From Dry to Wet Lab. *Open Microbiol. J.* **2017**, *11* (1), 330–338.

(7) Beier, M.; Hoheisel, J. D. Versatile Derivatisation of Solid Support Media for Covalent Bonding on DNA-Microchips. *Nucleic Acids Res.* **1999**, *27* (9), 1970–1977.

(8) Steel, A. B.; Herne, T. M.; Tarlov, M. J. Electrochemical Quantitation of DNA Immobilized on Gold. *Anal. Chem.* **1998**, *70* (22), 4670–4677.

(9) Xue, Y.; Li, X.; Li, H.; Zhang, W. Quantifying Thiol-Gold Interactions towards the Efficient Strength Control. *Nat. Commun.* **2014**, *5*.

(10) Gajos, K.; Petrou, P.; Budkowski, A.; Awsiuk, K.; Bernasik, A.; Misiakos, K.; Rysz, J.; Raptis, I.; Kakabakos, S. Imaging and Spectroscopic Comparison of Multi-Step Methods to Form DNA Arrays Based on the Biotin-Streptavidin System. *Analyst* **2015**, *140* (4), 1127–1139.

(11) Sabanayagam, C. R. Oligonucleotide Immobilization on Micropatterned Streptavidin Surfaces. *Nucleic Acids Res.* **2000**, *28* (8), 33e – 0.

(12) Wei, C. W.; Cheng, J. Y.; Huang, C. T.; Yen, M. H.; Young, T. H. Using a Microfluidic Device for 1 Ml DNA Microarray Hybridization in 500 S. *Nucleic Acids Res.* **2005**, *33* (8), 1–11.

(13) Qiao, W.; Chiang, H. C.; Xie, H.; Levicky, R. Surface vs. Solution Hybridization: Effects of Salt, Temperature, and Probe Type. *Chem. Commun.* **2015**, *51* (97), 17245–17248.

(14) Brown, R. W.; Schena, M.; Shalon, D.; Brown, P. Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray. *Science* **1995**,

270, 467–470.

(15) Shi, S. J. DNA Exhibits Multi-Stranded Binding Recognition on Glass Microarrays. *Nucleic Acids Res.* **2001**, *29* (20), 4251–4256.

(16) Wang, X.; Lim, H. J.; Son, A. Characterization of Denaturation and Renaturation of DNA for DNA Hybridization. *Environ. Health Toxicol.* **2014**, *29*, e2014007.

(17) Sinigaglia, C.; Thiel, D.; Hejnol, A.; Houliston, E.; Leclère, L. A Safer, Urea-Based in Situ Hybridization Method Improves Detection of Gene Expression in Diverse Animal Species. *Dev. Biol.* **2017**, *434*, 15–23.

(18) Sedighi, A.; Li, P. C. H. Kras Gene Codon 12 Mutation Detection Enabled by Gold Nanoparticles Conducted in a Nanobioarray Chip. *Anal. Biochem.* **2014**, *448* (1), 58–64.

(19) Sedighi, A.; Li, P. C. H.; Pekcevik, I. C.; Gates, B. D. A Proposed Mechanism of the Influence of Gold Nanoparticles on DNA Hybridization. *ACS Nano* **2014**, *8* (7), 6765–6777.

(20) Jaganade, T.; Chattopadhyay, A.; Pazhayam, N. M.; Priyakumar, U. D. Energetic, Structural and Dynamic Properties of Nucleobase-Urea Interactions That Aid in Urea Assisted RNA Unfolding. *Sci. Rep.* **2019**, *9* (1).

(21) Wang, L.; Li, P. C. H. Gold Nanoparticle-Assisted Single Base-Pair Mismatch Discrimination on a Microfluidic Microarray Device. *Biomicrofluidics* **2010**, *4* (3).

(22) Huang, S. H.; Chang, Y. S.; Juang, J. M. J.; Chang, K. W.; Tsai, M. H.; Lu, T. P.; Lai, L. C.; Chuang, E. Y.; Huang, N. T. An Automated Microfluidic DNA Microarray Platform for Genetic Variant Detection in Inherited Arrhythmic Diseases. *Analyst* **2018**, *143* (6), 1367–1377.

(23) Wang, L.; Li, P. C. H. Microfluidic DNA Microarray Analysis: A Review. *Anal. Chim. Acta* **2011**, *687* (1), 12–27.

(24) Ravan, H.; Kashanian, S.; Sanadgol, N.; Badoei-Dalfard, A.; Karami, Z. Strategies for Optimizing DNA Hybridization on Surfaces. *Anal. Biochem.* **2014**, *444* (1), 41–46.

(25) Weng, X.; Jiang, H.; Li, D. Microfluidic DNA Hybridization Assays. *Microfluid. Nanofluidics* **2011**, *11* (4), 367–383.

(26) Campuzano, S.; Yáñez-Sedeño, P.; Pingarrón, J. M. Tailoring Sensitivity in Electrochemical Nucleic Acid Hybridization Biosensing: Role of Surface Chemistry and Labeling Strategies. *ChemElectroChem* **2019**, *6* (1), 60–72.

(27) Zhang, X.; Hu, H. DNA Molecules Site-Specific Immobilization and Their Applications. *Cent. Eur. J. Chem.* **2014**, *12* (10), 977–993.

(28) SantaLucia Jr., J.; Allawi, H. T. Thermodynamics and NMR of Internal G.T Mismatches in DNA. *Biochemistry* **1997**, *36* (34), 10581–10594.

(29) M. Zuker. Mfold Web Server for Nucleic Acid Folding and Hybridization

Prediction. *Nucleic Acids Res.* **2003**, *31* (13), 3406–3415.

(30) Kibbe, W. A. OligoCalc: An Online Oligonucleotide Properties Calculator. *Nucleic Acids Res.* **2007**, *35* (SUPPL.2).

(31) Weber, G. Optimization Method for Obtaining Nearest-Neighbour DNA Entropies and Enthalpies Directly from Melting Temperatures. *Bioinformatics* **2015**, *31* (6), 871–877.

(32) Hooyberghs, J.; Van Hummelen, P.; Carlon, E. The Effects of Mismatches on Hybridization in DNA Microarrays: Determination of Nearest Neighbor Parameters. *Nucleic Acids Res.* **2009**, *37* (7).

(33) Hadiwikarta, W. W.; Walter, J. C.; Hooyberghs, J.; Carlon, E. Probing Hybridization Parameters from Microarray Experiments: Nearest-Neighbor Model and Beyond. *Nucleic Acids Res.* **2012**, *40* (18).

(34) Tinoco, I. A Unified View of Polymer , Dumbbell , and Oligonucleotide DNA Nearest-Neighbor Thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1460–1465.

(35) SantaLucia, J.; Hicks, D. The Thermodynamics of DNA Structural Motifs. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415–440.

(36) Zhang, J.; Fang, J.; Duan, W.; Wu, L.; Zhang, A.; Dalchau, N.; Yordanov, B.; Petersen, R.; Phillips, A.; Zhang, D. Y. Predicting DNA Hybridization Kinetics from Sequence. *Nat. Chem.* **2017**, *10*, 91–98.

(37) Barone, V.; Cacelli, I.; Ferretti, A.; Monti, S.; Prampolini, G. Sensors for DNA Detection: Theoretical Investigation of the Conformational Properties of Immobilized Single-Strand DNA. *Phys. Chem. Chem. Phys.* **2009**, *11* (45), 10644–10656.

(38) Wang, S.; Sun, Y.; Gan, W.; Liu, Y.; Xiang, G.; Wang, D.; Wang, L.; Cheng, J.; Liu, P. An Automated Microfluidic System for Single-Stranded DNA Preparation and Magnetic Bead-Based Microarray Analysis. *Biomicrofluidics* **2015**, *9* (2).

(39) Lin, F.; Zhou, Y.; Li, Q.; Zhou, X.; Shao, Y.; Habermeyer, B.; Wang, H.; Shi, X.; Xu, Z. Prototropically Allosteric Probe for Superbly Selective DNA Analysis. *Anal. Chem.* **2017**, *89* (17), 9299–9306.

(40) Deng, Z.; Ge, Y.; Cao, Q.; Han, K. The Detection of a Transgenic Soybean Biochip Using Gold Label Silver Stain Technology. *Bioorganic Med. Chem. Lett.* **2011**, *21* (22), 6905–6908.

(41) Krjutškov, K.; Andreson, R.; Mägi, R.; Nikopensius, T.; Khrunin, A.; Mihailov, E.; Tammekivi, V.; Sork, H.; Remm, M.; Metspalu, A. Development of a Single Tube 640-Plex Genotyping Method for Detection of Nucleic Acid Variations on Microarrays. *Nucleic Acids Res.* **2008**, *36* (12).

(42) Wang, L.; Li, P. C. H. Flexible Microarray Construction and Fast DNA Hybridization Conducted on a Microfluidic Chip for Greenhouse Plant Fungal

Pathogen Detection. *J. Agric. Food Chem.* **2007**, *55*, 10509–10516.

(43) Behjati, S.; Tarpey, P. S. What Is next Generation Sequencing? *Arch. Dis. Child. Educ. Pract. Ed.* **2013**, *98* (6), 236–238.

(44) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74* (12), 5463–5467.

(45) Smith, L. M.; Sanders, J. Z.; Kaiser, R. J.; Hughes, P.; Dodd, C.; Connell, C. R.; Heiner, C.; Kent, S. B. H.; Hood, L. E. Fluorescence Detection in Automated DNA Sequence Analysis. *Nature* **1986**, *321* (6071), 674–679.

(46) Schmutz, J.; Wheeler, J.; Grimwood, J.; Dickson, M.; Yang, J.; Caolle, C.; Bajorek, E.; Black, S.; Chan, Y. M.; Denys, M.; et al. Quality Assessment of the Human Genome Sequence. *Nature* **2004**, *429* (6990), 365–368.

(47) Nyrén, P.; Pettersson, B.; Uhlén, M. Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay. *Anal. Biochem.* **1993**, *208* (1), 171–175.

(48) Ronaghi, M.; Uhlén, M.; Nyrén, P. A Sequencing Method Based on Real-Time Pyrophosphate. *Science (80-.).* **1998**, *281* (5375), 363–365.

(49) Purushothaman, S.; Toumazou, C.; Ou, C. P. Protons and Single Nucleotide Polymorphism Detection: A Simple Use for the Ion Sensitive Field Effect Transistor. *Sensors Actuators, B Chem.* **2006**, *114* (2), 964–968.

(50) Pennisi, E. Semiconductors Inspire New Sequencing Technologies. *Science (80-.).* **2010**, *327* (5970), 1190.

(51) Canard, B.; Sarfati, R. S. DNA Polymerase Fluorescent Substrates with Reversible 3′-Tags. *Gene* **1994**, *148* (1), 1–6.

(52) Huang, J.; Liang, X.; Xuan, Y.; Geng, C.; Li, Y.; Lu, H.; Qu, S.; Mei, X.; Chen, H.; Yu, T.; et al. A Reference Human Genome Dataset of the BGISEQ-500 Sequencer. *Gigascience* **2017**, *6* (5), 1–9.

(53) Deamer, D. W. Characterization of Individual Polynucleotide Molecules Using a Membrane Channel. *Chemtracts* **1997**, *10* (3), 255–257.

(54) Stoddart, D.; Heron, A. J.; Mikhailova, E.; Maglia, G.; Bayley, H. Single-Nucleotide Discrimination in Immobilized DNA Oligonucleotides with a Biological Nanopore. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (19), 7702–7707.

(55) Soni, G. V.; Singer, A.; Yu, Z.; Sun, Y.; McNally, B.; Meller, A. Synchronous Optical and Electrical Detection of Biomolecules Traversing through Solid-State Nanopores. *Rev. Sci. Instrum.* **2010**, *81* (1).

(56) Simpson, J.; Workman, R.; David, M.; Timp, W. Detecting DNA Methylation Using the Oxford Nanopore Technologies MinION Sequencer. *Nat. Methods* **2016**, 047142.

(57) Ashton, P. M.; Nair, S.; Dallman, T.; Rubino, S.; Rabsch, W.; Mwaigwisya, S.;

Wain, J.; O'Grady, J. MinION Nanopore Sequencing Identifies the Position and Structure of a Bacterial Antibiotic Resistance Island. *Nat. Biotechnol.* **2015**, *33* (3), 296–302.

(58) Taylor, D. A. *Ginseng, the Divine Root*; Algonquin Books of Chapel Hill: New York, 2006.

(59) Coon, J. T.; Ernst, E. Panax Ginseng. *Drug Saf.* **2002**, *25*, 323–344.

(60) Baeg, I. H.; So, S. H. The World Ginseng Market and the Ginseng (Korea). *J. Ginseng Res.* **2013**, *37* (1), 1–7.

(61) Waminal, N. E.; Park, H. M.; Ryu, K. B.; Kim, J. H.; Yang, T. J.; Kim, H. H. Karyotype Analysis of Panax Ginseng C.A.Meyer, 1843 (Araliaceae) Based on RDNA Loci and DAPI Band Distribution. *Comp. Cytogenet.* **2012**, *6* (4), 425–441.

(62) G. Obae, S. Nuclear DNA Content and Genome Size of American Ginseng. *J. Med. Plants Res.* **2012**, *6* (32).

(63) Kim, Y. J.; Silva, J.; Zhang, D.; Shi, J.; Joo, S. C.; Jang, M. G.; Kwon, W. S.; Yang, D. C. Development of Interspecies Hybrids to Increase Ginseng Biomass and Ginsenoside Yield. *Plant Cell Rep.* **2016**, *35* (4), 779–790.

(64) Seo, K. H.; Mohanan, P.; Yang, D. U.; Kim, Y. J.; Kwon, W. S.; Yang, D. C. Interspecies Hybrids of Panax Ginseng Meyer New Line 0837 and Panax Quinquefolius Generated Superior F1 Hybrids with Greater Biomass and Ginsenoside Contents. *Hortic. Environ. Biotechnol.* **2019**, *60* (4), 573–583.

(65) Tansakul, P.; Shibuya, M.; Kushiro, T.; Ebizuka, Y. Dammarenediol-II Synthase, the First Dedicated Enzyme for Ginsenoside Biosynthesis, in Panax Ginseng. *FEBS Lett.* **2006**, *580* (22), 5143–5149.

(66) Han, J. Y.; Kim, H. J.; Kwon, Y. S.; Choi, Y. E. The Cyt P450 Enzyme CYP716A47 Catalyzes the Formation of Protopanaxadiol from Dammarenediol-II during Ginsenoside Biosynthesis in Panax Ginseng. *Plant Cell Physiol.* **2011**, *52* (12), 2062–2073.

(67) Li, J.; Li, H.; Liu, D.; Liu, S.; Li, J.; Wang, J. Analysis of Ginsenoside Content, Functional Genes Involved in Ginsenosides Biosynthesis, and Activities of Antioxidant Enzymes in Panax Quinquefolium L. Adventitious Roots by Fungal Elicitors. *Res. Chem. Intermed.* **2017**, *43* (4), 2415–2432.

(68) Liu, C. Y.; Zhou, R. X.; Sun, C. K.; Jin, Y. H.; Yu, H. S.; Zhang, T. Y.; Xu, L. Q.; Jin, F. X. Preparation of Minor Ginsenosides C-Mc, C-Y, F2, and C-K from American Ginseng PPD-Ginsenoside Using Special Ginsenosidase Type-i from Aspergillus Niger g.848. *J. Ginseng Res.* **2015**, *39* (3), 221–229.

(69) Cui, C. H.; Liu, Q. M.; Kim, J. K.; Sung, B. H.; Kim, S. G.; Kim, S. C.; Im, W. T. Identification and Characterization of a Mucilaginibacter Sp. Strain QM49 β-Glucosidase and Its Use in the Production of the Pharmaceutically Active Minor Ginsenosides (S)-Rh1 and (S)-Rg2. *Appl. Environ. Microbiol.* **2013**, *79* (19), 5788–5798.

(70) Uhr, L.; Chen, Y.; Sit, D.; Li, P. C. H. Ginsenosides in Commercial Ginseng Products Analyzed by Liquid Chromatography-Tandem Mass Spectrometry. *ISRN Anal. Chem.* **2014**, *2014*, 1–8.

(71) Tam, D. N. H.; Truong, D. H.; Nguyen, T. T. H.; Quynh, L. N.; Tran, L.; Nguyen, H. D.; Shamandy, B. E.; Le, T. M. H.; Tran, D. K.; Sayed, D.; et al. Ginsenoside Rh1: A Systematic Review of Its Pharmacological Properties. *Planta Med.* **2018**, *84* (3), 139–152.

(72) Liu, H.; Liu, M.; Jin, Z.; Yaqoob, S.; Zheng, M.; Cai, D.; Liu, J.; Guo, S. Ginsenoside Rg2 Inhibits Adipogenesis in 3T3-L1 Preadipocytes and Suppresses Obesity in High-Fat-Diet-Induced Obese Mice through the AMPK Pathway. *Food Funct.* **2019**, *10* (6), 3603–3614.

(73) Chen, C.-F.; Chiou, W.-F.; Zhang, J. Comparison of the Pharmacological Effects of Panax Ginseng and Panax Quinquefolium. *Acta Pharmacol. Sin.* **2008**, *29*, 1103–1108.

(74) Yang, W.; Qiao, X.; Li, K.; Fan, J.; Bo, T.; Guo, D. A.; Ye, M. Identification and Differentiation of Panax Ginseng, Panax Quinquefolium, and Panax Notoginseng by Monitoring Multiple Diagnostic Chemical Markers. *Acta Pharm. Sin. B* **2016**, *6* (6), 568–575.

(75) Chang, Y.; Huang, W. J.; Tien, L. T.; Wang, S. J. Ginsenosides Rg1 and Rb1 Enhance Glutamate Release through Activation of Protein Kinase A in Rat Cerebrocortical Nerve Terminals (Synaptosomes). *Eur. J. Pharmacol.* **2008**, *578* (1), 28–36.

(76) Yun, T. K.; Lee, Y. S.; Lee, Y. H.; Kim, S. I.; Yun, H. Y. Anticarcinogenic Effect of Panax Ginseng C.A. Meyer and Identification of Active Compounds. *J. Korean Med. Sci.* **2001**, *16 Suppl*.

(77) Popovich, D. G.; Kitts, D. D. Structure-Function Relationship Exists for Ginsenosides in Reducing Cell Proliferation and Inducing Apoptosis in the Human Leukemia (THP-1) Cell Line. *Arch. Biochem. Biophys.* **2002**, *406* (1), 1–8.

(78) Yuan, C. S.; Wei, G.; Dey, L.; Karrison, T.; Nahlik, L.; Maleckar, S.; Kasza, K.; Ang-Lee, M.; Moss, J. Brief Communication: American Ginseng Reduces Warfarin's Effect in Healthy Patients. A Randomized, Controlled Trial. *Ann. Intern. Med.* **2004**, *141* (1).

(79) Lee, S. H.; Ahn, Y. M.; Ahn, S. Y.; Doo, H. K.; Lee, B. C. Interaction between Warfarin and Panax Ginseng in Ischemic Stroke Patients. *J. Altern. Complement. Med.* **2008**, *14* (6), 715–721.

(80) Lee, Y. H.; Lee, B. K.; Choi, Y. J.; Yoon, I. K.; Chang, B. C.; Gwak, H. S. Interaction between Warfarin and Korean Red Ginseng in Patients with Cardiac Valve Replacement. *Int. J. Cardiol.* **2010**, *145* (2), 275–276.

(81) Lin, J. F.; Fan, L. L.; Li, B. W.; Zhao, R. R.; Jiang, L. G.; Zhang, B. C.; Lu, Y. S.; Shao, J. W. A Study to Evaluate Herb-Drug Interaction Underlying Mechanisms:

An Investigation of Ginsenosides Attenuating the Effect of Warfarin on Cardiovascular Diseases. *Eur. J. Pharm. Sci.* **2020**, *142*.

(82) Zhang, B.; Zhou, W. J.; Gu, C. J.; Wu, K.; Yang, H. L.; Mei, J.; Yu, J. J.; Hou, X. F.; Sun, J. S.; Xu, F. Y.; et al. The Ginsenoside PPD Exerts Anti-Endometriosis Effects by Suppressing Estrogen Receptor-Mediated Inhibition of Endometrial Stromal Cell Autophagy and NK Cell Cytotoxicity Article. *Cell Death Dis.* **2018**, *9* (5).

(83) Yang, L.; Zhang, X. ying; Li, K.; Li, A. ping; Yang, W. dong; Yang, R.; Wang, P.; Zhao, Z. han; Cui, F.; Qin, Y.; et al. Protopanaxadiol Inhibits Epithelial–Mesenchymal Transition of Hepatocellular Carcinoma by Targeting STAT3 Pathway. *Cell Death Dis.* **2019**, *10* (9).

(84) Xu, F. Y.; Shang, W. Q.; Yu, J. J.; Sun, Q.; Li, M. Q.; Sun, J. S. The Antitumor Activity Study of Ginsenosides and Metabolites in Lung Cancer Cell. *Am. J. Transl. Res.* **2016**, *8* (4), 1708–1718.

(85) Wang, H. P.; Zhang, Y. B.; Yang, X. W.; Zhao, D. Q.; Wang, Y. P. Rapid Characterization of Ginsenosides in the Roots and Rhizomes of Panax Ginseng by UPLC-DAD-QTOF-MS/MS and Simultaneous Determination of 19 Ginsenosides by HPLC-ESI-MS. *J. Ginseng Res.* **2016**, *40* (4), 382–394.

(86) Zhou, S. S.; Xu, J.; Kong, M.; Yip, K. M.; Xu, J. Di; Shen, H.; Zhao, Z. Z.; Li, S. L.; Chen, H. B. Synchronous Characterization of Carbohydrates and Ginsenosides Yields Deeper Insights into the Processing Chemistry of Ginseng. *J. Pharm. Biomed. Anal.* **2017**, *145*, 59–70.

(87) Tian, Y.; Lu, Y.; Xie, J.; Cheng, Y.; Qi, R.; Wu, Y.; Zhang, S. Rapid Determination of Ginsenoside Rg1, Re and Rb1 in Ginseng Samples by Capillary Electrophoresis. *Anal. Methods* **2009**, *1* (3), 203–207.

(88) Jee, H. S.; Chang, K. H.; Park, S. H.; Kim, K. T.; Paik, H. D. Morphological Characterization, Chemical Components, and Biofunctional Activities of Panax Ginseng, Panax Quinquefolium, and Panax Notoginseng Roots: A Comparative Study. *Food Rev. Int.* **2014**, *30* (2), 91–111.

(89) Kim, H. M.; Song, Y.; Hyun, G. H.; Long, N. P.; Park, J. H.; Hsieh, Y. S. Y.; Kwon, S. W. Characterization and Antioxidant Activity Determination of Neutral and Acidic Polysaccharides from Panax Ginseng C. A. Meyer. *Molecules* **2020**, *25* (4).

(90) Liu, Z.; Wang, C. Z.; Zhu, X. Y.; Wan, J. Y.; Zhang, J.; Li, W.; Ruan, C. C.; Yuan, C. S. Dynamic Changes in Neutral and Acidic Ginsenosides with Different Cultivation Ages and Harvest Seasons: Identification of Chemical Characteristics for Panax Ginseng Quality Control. *Molecules* **2017**, *22* (5).

(91) Harnly, J.; Chen, P.; Harrington, P. D. B. Probability of Identification: Adulteration of American Ginseng with Asian Ginseng. *J. AOAC Int.* **2013**, *96* (6), 1258–1265.

(92) Zhang, J. J.; Su, H.; Zhang, L.; Liao, B. S.; Xiao, S. M.; Dong, L. L.; Hu, Z. G.; Wang, P.; Li, X. W.; Huang, Z. H.; et al. Comprehensive Characterization for Ginsenosides Biosynthesis in Ginseng Root by Integration Analysis of Chemical and Transcriptome. *Molecules* **2017**, *22* (6).

(93) Choi, H. Il; Waminal, N. E.; Park, H. M.; Kim, N. H.; Choi, B. S.; Park, M.; Choi, D.; Lim, Y. P.; Kwon, S. J.; Park, B. S.; et al. Major Repeat Components Covering One-Third of the Ginseng (Panax Ginseng C.A. Meyer) Genome and Evidence for Allotetraploidy. *Plant J.* **2014**, *77* (6), 906–916.

(94) Kim, K.; Lee, S. C.; Lee, J.; Lee, H. O.; Joh, H. J.; Kim, N. H.; Park, H. S.; Yang, T. J. Comprehensive Survey of Genetic Diversity in Chloroplast Genomes and 45S NrDNAs within Panax Ginseng Species. *PLoS One* **2015**, *10* (6).

(95) Kim, J. H.; Kim, M. K.; Wang, H.; Lee, H. N.; Jin, C. G.; Kwon, W. S.; Yang, D. C. Discrimination of Korean Ginseng (Panax Ginseng Meyer) Cultivar Chunpoong and American Ginseng (Panax Quinquefolius) Using the Auxin Repressed Protein Gene. *J. Ginseng Res.* **2016**, *40* (4), 395–399.

(96) Wang, H.; Wang, J.; Li, G. A Simple Real-Time Polymerase Chain Reaction (PCR)-Based Assay for Authentication of the Chinese Panax Ginseng Cultivar Damaya from a Local Ginseng Population. *Genet. Mol. Res.* **2016**, *15* (2), 1–10.

(97) Hu, W.; Liu, N.; Tian, Y.; Zhang, L. Molecular Cloning, Expression, Purification, and Functional Characterization of Dammarenediol Synthase from Panax Ginseng. *Biomed Res. Int.* **2012**, *2013*, 1–7.

(98) Wang, L.; Zhao, S.-J.; Cao, H.; Sun, Y. The Isolation and Characterization of Dammarenediol Synthase Gene from Panax Quinquefolius and Its Heterologous Co-Expression with Cytochrome P450 Gene PqD12H in Yeast. *Funct. Integr. Genomics* **2014**, *14*, 545–557.

(99) Zhou, H.; Wu, W.; Liu, L.; Cheng, C. Method for Molecular Authentication of Panax Ginseng and Panax Quinquefolius Using SNP Markers in Dammarenediol Synthase (DS) Gene. *Pat. Appl. AU2015100441*.

(100) Qin, J.; Leung, F. C.; Fung, Y.; Zhu, D.; Lin, B. Rapid Authentication of Ginseng Species Using Microchip Electrophoresis with Laser-Induced Fluorescence Detection. *Anal. Bioanal. Chem.* **2005**, *381* (4), 812–819.

(101) Um, J. Y.; Chung, H. S.; Kim, M. S.; Na, H. J.; Kwon, H. J.; Kim, J. J.; Lee, K. M.; Lee, S. J.; Lim, J. P.; Do, K. R.; et al. Molecular Authentication of Panax Ginseng Species by RAPD Analysis and PCR-RFLP. *Biol. Pharm. Bull.* **2001**, *24* (8), 872–875.

(102) Lu, K. T.; Lee, H. C.; Liu, F. S.; Lo, C. F.; Lin, J. H. Identification of Ginseng Radix in Chinese Medicine Preparations by Nested PCR-DNA Sequencing Method and Nested PCR-Restriction Fragment Length Polymorphism. *J. Food Drug Anal.* **2010**, *18* (1), 58–63.

(103) Ha, W. Y.; Shaw, P. C.; Liu, J.; Yau, F. C. F.; Wang, J. Authentication of Panax

Ginseng and Panax Quinquefolius Using Amplified Fragment Length Polymorphism (AFLP) and Directed Amlification of Minisatellite Region DNA (DAMD). *J. Agric. Food Chem.* **2002**, *50* (7), 1871–1875.

(104) Jo, I. H.; Bang, K. H.; Kim, Y. C.; Lee, J. W.; Seo, A. Y.; Seong, B. J.; Kim, H. H.; Kim, D. H.; Cha, S. W.; Cho, Y. G.; et al. Rapid Identification of Ginseng Cultivars (Panax Ginseng Meyer) Using Novel SNP-Based Probes. *J. Ginseng Res.* **2011**, *35* (4), 504–513.

(105) Sedighi, A.; Oberc, C.; Whitehall, V.; Li, P. C. H. NanoHDA: A Nanoparticle-Assisted Isothermal Amplification Technique for Genotyping Assays. *Nano Res.* **2017**, *10* (1), 12–21.

(106) An, L.; Tang, W.; Ranalli, T. A.; Kim, H. J.; Wytiaz, J.; Kong, H. Characterization of a Thermostable UvrD Helicase and Its Participation in Helicase-Dependent Amplification. *J. Biol. Chem.* **2005**, *280* (32), 28952–28958.

(107) Kausar, A.; McKay, R. D.; Lam, J.; Bhogal, R. S.; Tang, A. Y.; Gibbs-Davis, J. M. Tuning DNA Stability to Achieve Turnover in Template for an Enzymatic Ligation Reaction. *Angew. Chemie - Int. Ed.* **2011**, *50* (38), 8922–8926.

(108) Kausar, A.; Mitran, C. J.; Li, Y.; Gibbs-Davis, J. M. Rapid, Isothermal DNA Self-Replication Induced by a Destabilizing Lesion. *Angew. Chemie - Int. Ed.* **2013**, *52* (40), 10577–10581.

(109) Shin, Y.; Kim, J.; Lee, T. Y. A Solid Phase-Bridge Based DNA Amplification Technique with Fluorescence Signal Enhancement for Detection of Cancer Biomarkers. *Sensors Actuators, B Chem.* **2014**, *199*, 220–225.

(110) Ikuta, S.; Takagi, K.; Wallace, R. B.; Itakura, K. Dissociation Kinetics of 19 Base Paired Oligonucleotide-DNA Duplexes Containing Different Single Mismatched Base Pairs. *Nucleic Acids Res.* **1987**, *15* (2), 797–811.

(111) Sedighi, A.; Whitehall, V.; Li, P. C. H. Enhanced Destabilization of Mismatched DNA Using Gold Nanoparticles Offers Specificity without Compromising Sensitivity for Nucleic Acid Analyses. *Nano Res.* **2015**, *8* (12), 3922–3933.

(112) Zheng, Y.; Bergold, A.; Duffel, M. W. Affinity Labeling of Aryl Sulfotransferase IV. Identification of a Peptide Sequence at the Binding Site for 3'-Phosphoadenosine-5'-Phosphosulfate. *J. Biol. Chem.* **1994**, *269* (48), 30313–30319.

(113) Li, C.; Li, H.; Qin, J.; Lin, B. Rapid Discrimination of Single-Nucleotide Mismatches Based on Reciprocating Flow on a Compact Disc Microfluidic Device. *Electrophoresis* **2009**, *30* (24), 4270–4276.

(114) Simard, C.; Lemieux, R.; Coté, S. Urea Substitutes Toxic Formamide as Destabilizing Agent in Nucleic Acid Hybridizations with RNA Probes. *Electrophoresis* **2001**, *22* (13), 2679–2683.

(115) Wang, F.; Wang, L.; Briggs, C.; Sicinska, E.; Gaston, S. M.; Mamon, H.; Kulke, M. H.; Zamponi, R.; Loda, M.; Maher, E.; et al. DNA Degradation Test Predicts

Success in Whole-Genome Amplification from Diverse Clinical Samples. *J. Mol. Diagnostics* **2007**, *9* (4), 441–451.

(116) Wang, X.; Ying, S.; Wei, X.; Yuan, J. Development of a Gold Nanoparticle-Based Universal Oligonucleotide Microarray for Multiplex and Low-Cost Detection of Foodborne Pathogens. *Int. J. Food Microbiol.* **2017**, *253*, 66–74.

(117) Tanner, N. A.; Zhang, Y.; Evans, T. C. Simultaneous Multiple Target Detection in Real-Time Loop-Mediated Isothermal Amplification. *Biotechniques* **2012**, *53* (2), 81–89.

(118) Kusumoto-Matsuo, R.; Kanda, T.; Kukimoto, I. Rolling Circle Replication of Human Papillomavirus Type 16 DNA in Epithelial Cell Extracts. *Genes to Cells* **2011**, *16* (1), 23–33.

(119) Toley, B. J.; Covelli, I.; Belousov, Y.; Ramachandran, S.; Kline, E.; Scarr, N.; Vermeulen, N.; Mahoney, W.; Lutz, B. R.; Yager, P. Isothermal Strand Displacement Amplification (ISDA): A Rapid and Sensitive Method of Nucleic Acid Amplification for Point-of-Care Diagnosis. *Analyst* **2015**, *140* (22), 7540–7549.

(120) Vincent, M.; Xu, Y.; Kong, H. Helicase-Dependent Isothermal DNA Amplification. *EMBO Rep.* **2004**, *5* (8), 795–800.

(121) Kausar, A.; Osman, E. A.; Gadzikwa, T.; Gibbs-Davis, J. M. The Presence of a 5′-Abasic Lesion Enhances Discrimination of Single Nucleotide Polymorphisms While Inducing an Isothermal Ligase Chain Reaction. *Analyst* **2016**, *141* (14), 4272–4277.

(122) Kotera, N.; Granzhan, A.; Teulade-Fichou, M. P. Comparative Study of Affinity and Selectivity of Ligands Targeting Abasic and Mismatch Sites in DNA Using a Fluorescence-Melting Assay. *Biochimie* **2016**, *128–129*, 133–137.

(123) Öhrmalm, C.; Jobs, M.; Eriksson, R.; Golbob, S.; Elfaitouri, A.; Benachenhou, F.; Strømme, M.; Blomberg, J. Hybridization Properties of Long Nucleic Acid Probes for Detection of Variable Target Sequences, and Development of a Hybridization Prediction Algorithm. *Nucleic Acids Res.* **2010**, *38* (21).

(124) Hayashi, K.; Nakazawa, M.; Ishizaki, Y.; Obayashi, A. Influence of Monovalent Cations on the Activity of T4 DNA Ligase in the Presence of Polyethylene Glycol. *Nucleic Acids Res.* **1985**, *13* (9), 3261–3271.

(125) Tong, J.; Cao, W.; Barany, F. Biochemical Properties of a High Fidelity DNA Ligase from Thermus Species AK16D. *Nucleic Acids Res.* **1999**, *27* (3), 788–794.

(126) Heng, L.; Richard, D. Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform. *Bioinformatics* **2009**, *25* (14), 1754–1760.

(127) Robinson, J. T.; Thorvaldsdóttir, H.; Wenger, A. M.; Zehir, A.; Mesirov, J. P. Variant Review with the Integrative Genomics Viewer. *Cancer Res.* **2017**, *77* (21), e31–e34.

(128) Das, G.; Lyngdoh, R. H. D. Can Configuration of Solitary Wobble Base Pairs Determine the Specificity and Degeneracy of the Genetic Code? Clues from Molecular Orbital Modelling Studies. *J. Mol. Struct. THEOCHEM* **2008**, *851* (1–3), 319–334.

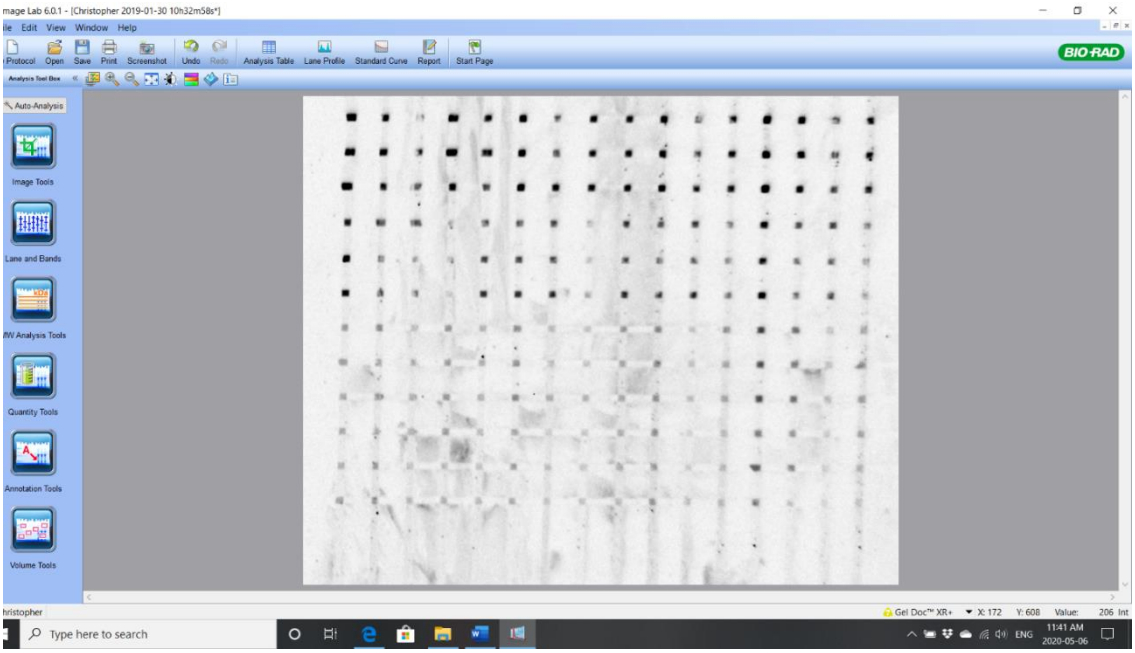
(129) Pritchard, C. E.; Southern, E. M. Effects of Base Mismatches on Joining of Short Oligodeoxynucleotides by DNA Ligases. *Nucleic Acids Res.* **1997**, *25* (17), 3403–3407.

Chapter 7 Appendices:

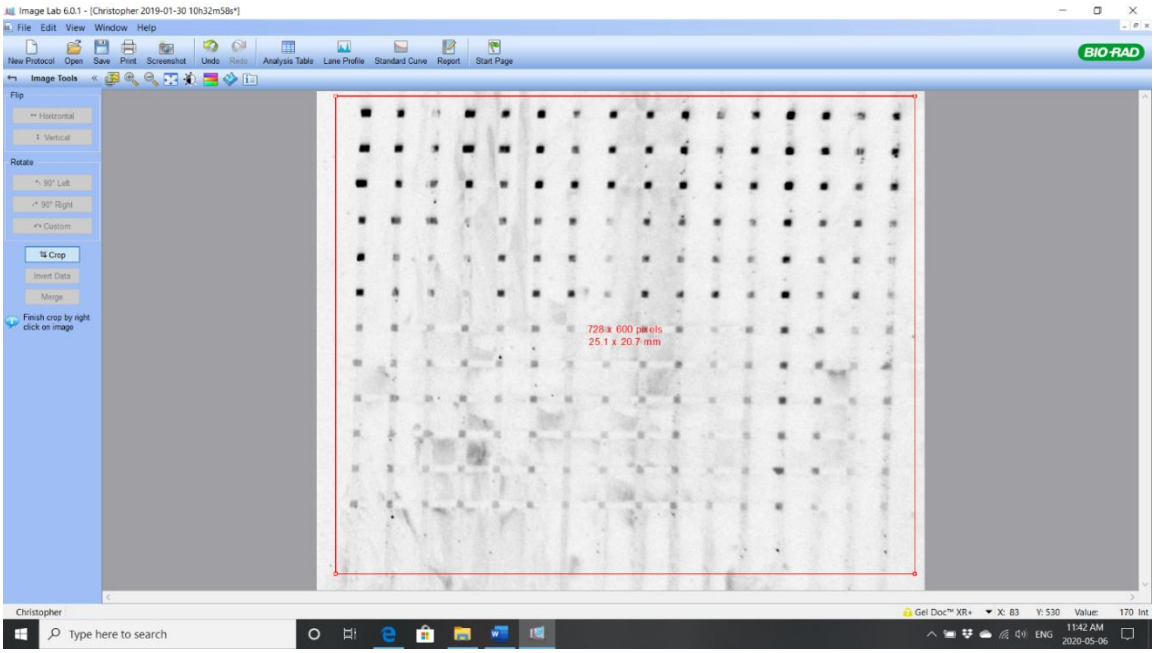
Appendix A: Image analysis using Image Lab

Images are collected from ChemiDoc Imager (BioRad), and they are analyzed using the software called Image Lab (6.0.1)

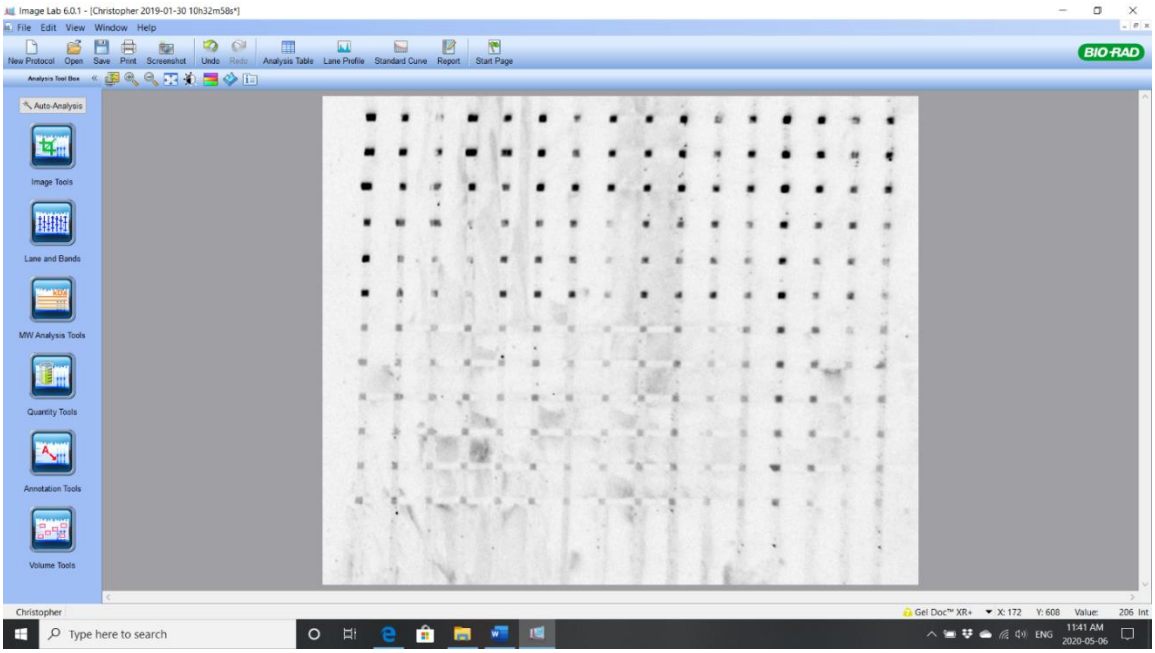
In the image analysis, fifteen steps are shown to identify fluorescence bands, determine pixel values and output them to a spreadsheet.



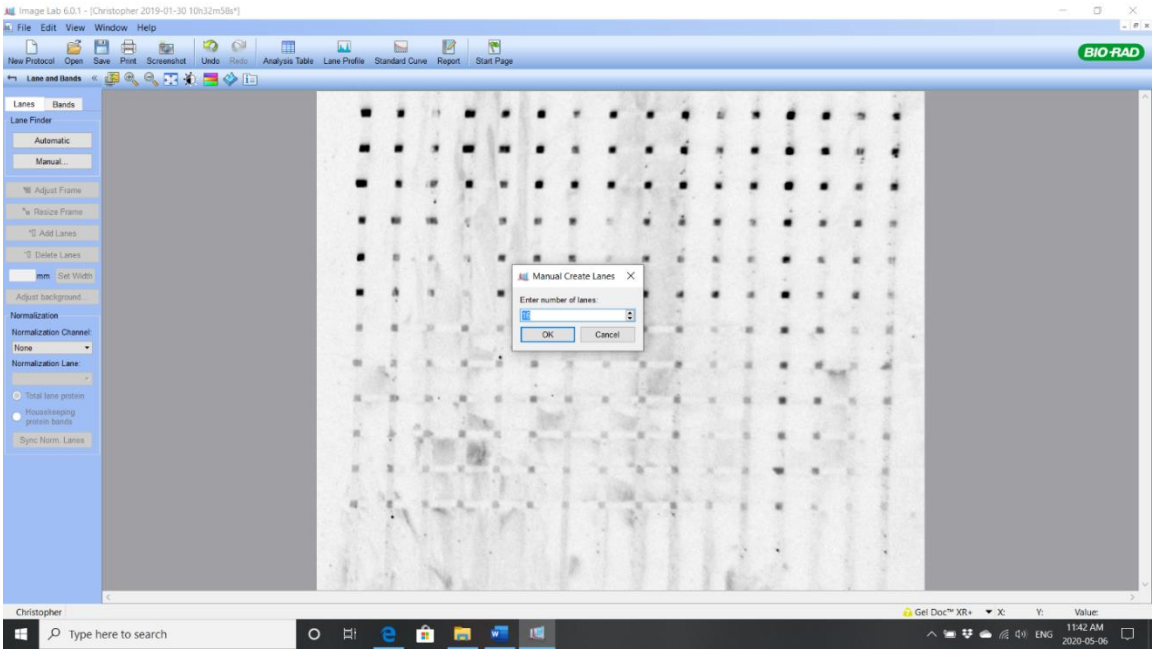
Step 1: On the left menu bar, select “Image Tools”.



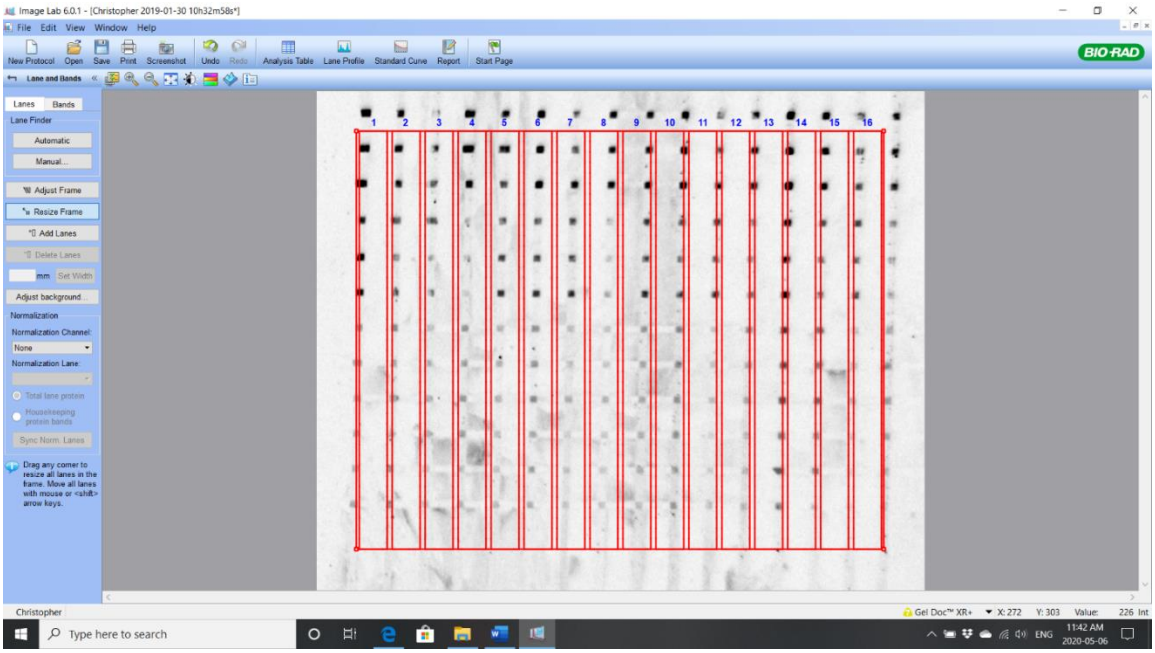
Step 2: The image was cropped down to the section that contains the microarray. Afterwards, the “Image Tools” menu was exited.



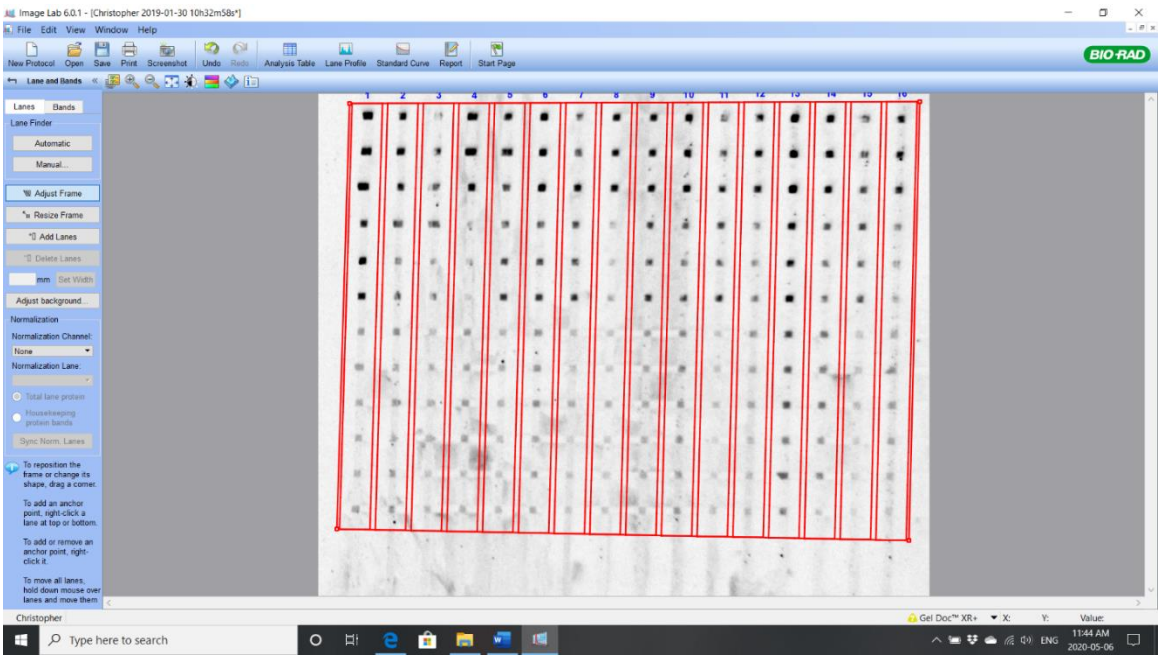
Step 3: On the left menu bar, select “Lanes and Bands” .



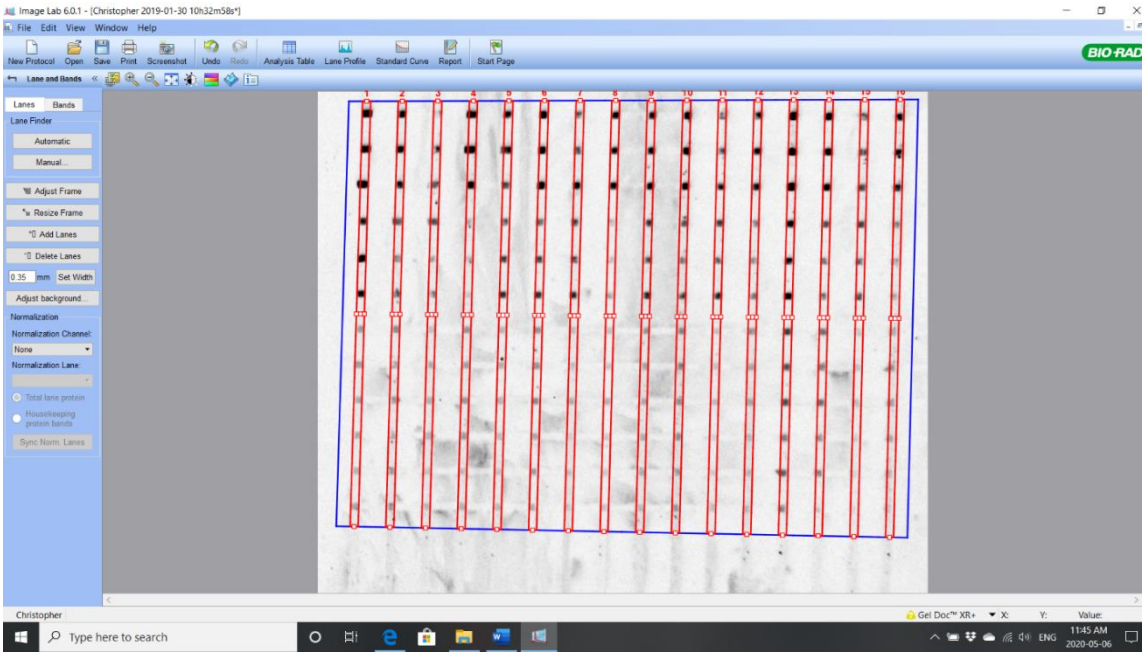
Step 4: While on the "Lanes" tab, select “Manual…” so that a popup window was opened. Enter the number of vertical lanes used (i.e. 16).



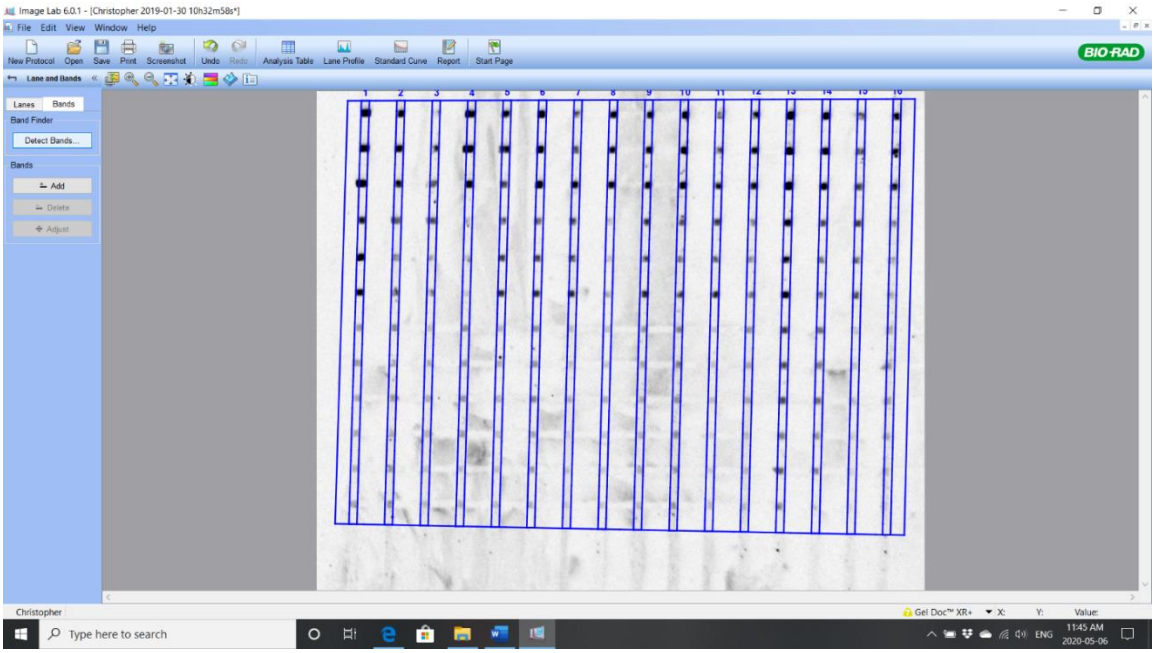
Step 5: The actions of Step 4 creates 16 vertical frames (red when selected) over the array.



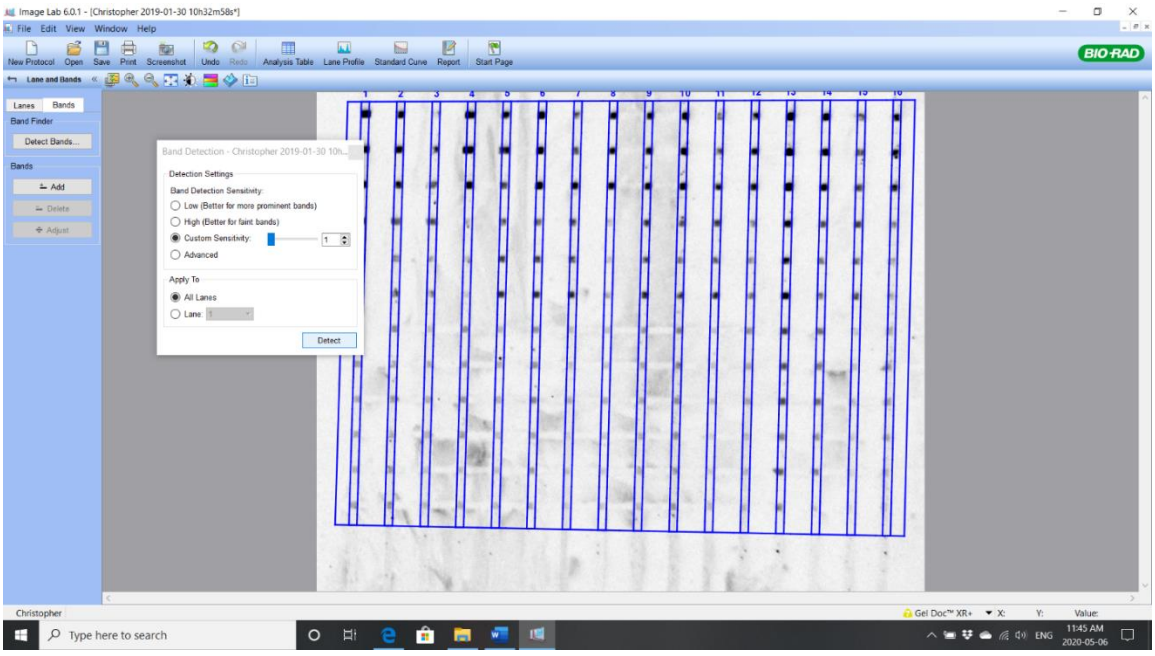
Step 6: Select the “Adjust Frame” option (it turns a light blue colour when selected) and the corners of the red frame were adjusted such that the vertical lanes were centered in the frames (indicated by the numbers 1 to 16).



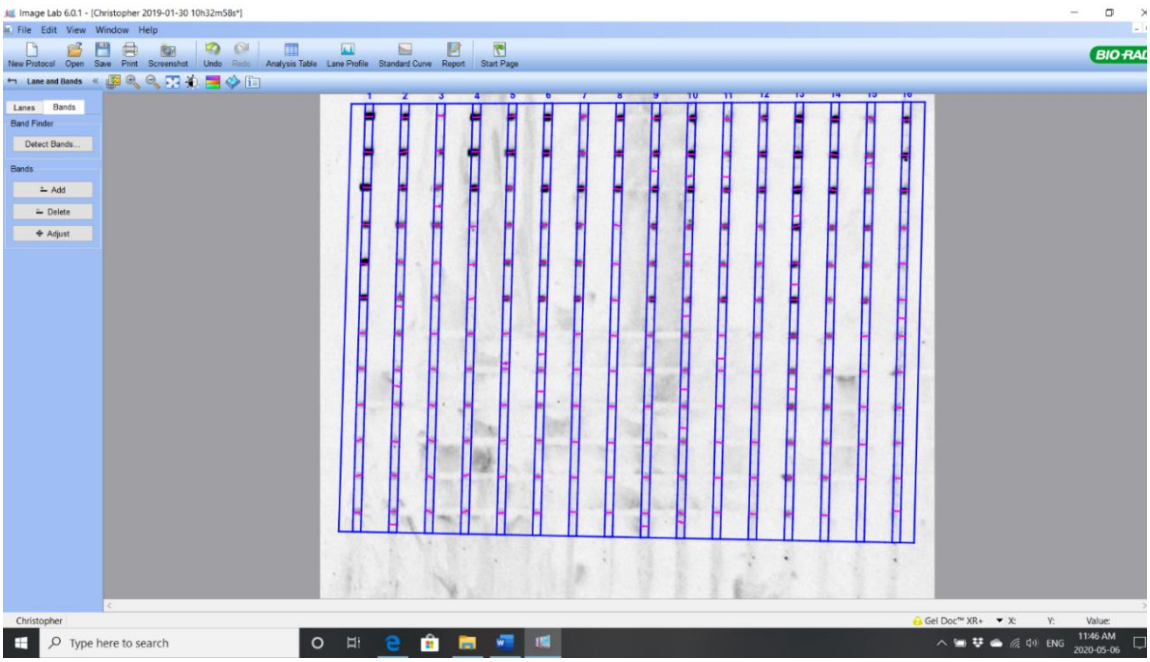
Step 7: The “Adjust Frame” option was deselected (turns from light blue to grey). Then the first and last lanes where selected while holding the Shift key (which selects all the lanes but not the perimeter of the frame). The lanes were then contracted using the adjustment points on them to match the width of the channels/spots.



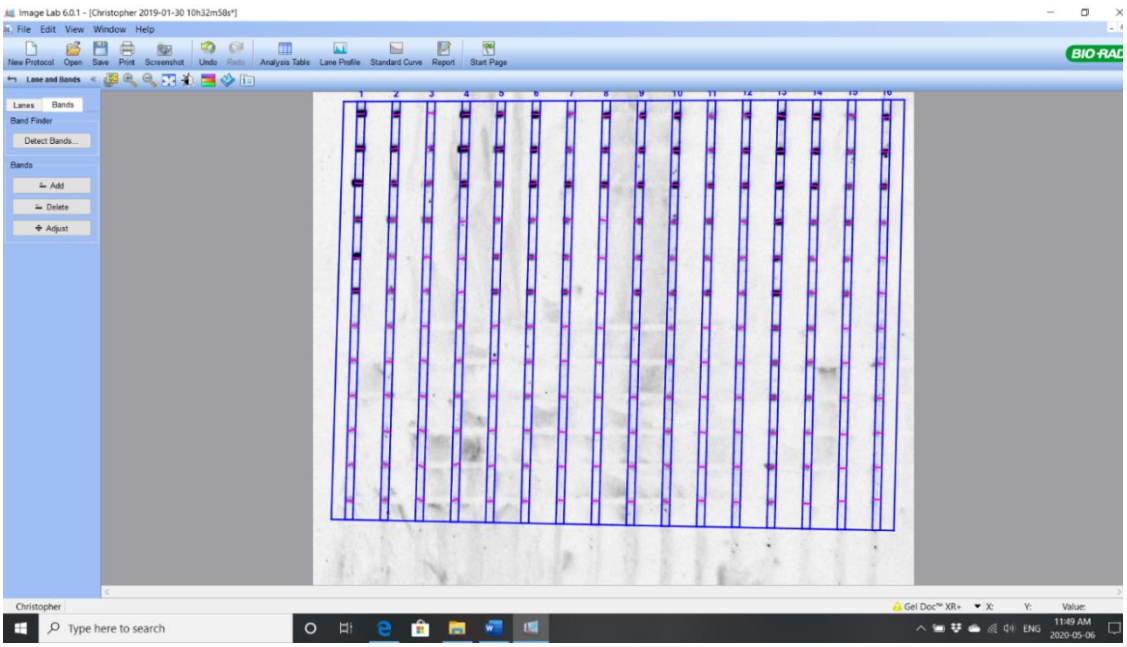
Step 8: Select the “Bands” tab.



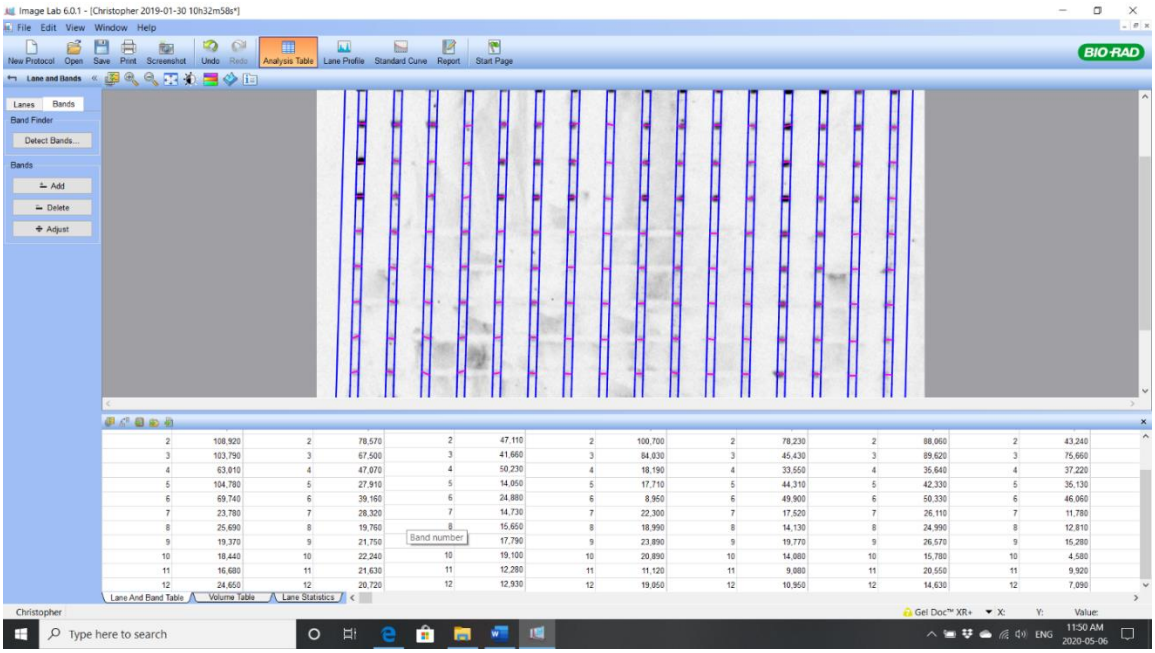
Step 9: In this menu, select “Detect Bands…” . In the popup window, press “Detect”. The options in the popup window were kept at their default settings.



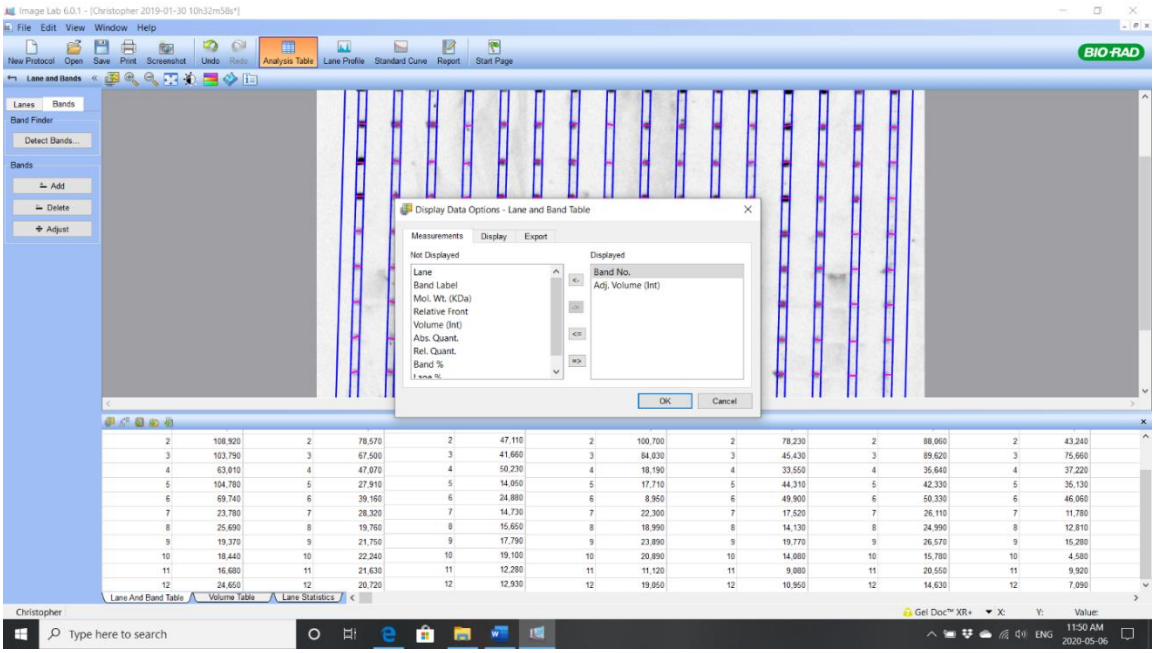
Step 10: The previous step 9 placed purple markers on all the parts of the lanes with significant fluorescence signal which ideally should have been all the intersections. However, as seen some intersections were not detected because their intensity was below the threshold for automatic detection while there were also false positives caused by small particles scattering light during the fluorescence scan. Both of these problems were resolved by using “Add” to insert missing bands and “Delete” to remove false positives. “Adjust” was also used to adjust the position of band.



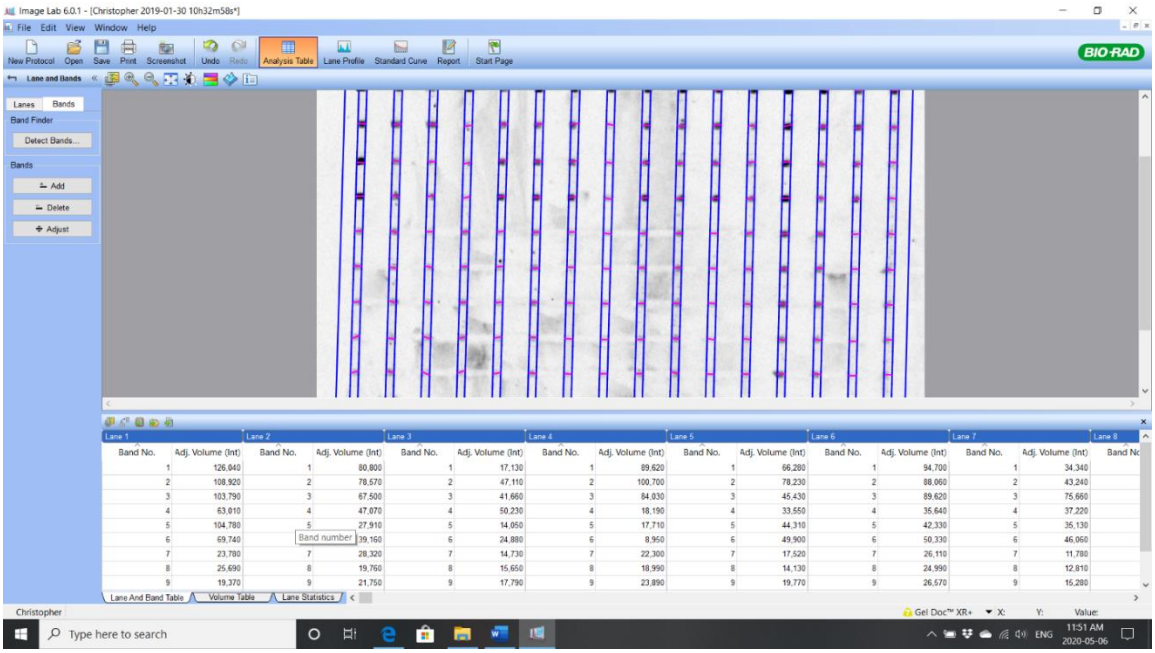
Step 11: After adjustment, this resulted in the grid arrangement of the bands (purple markers).



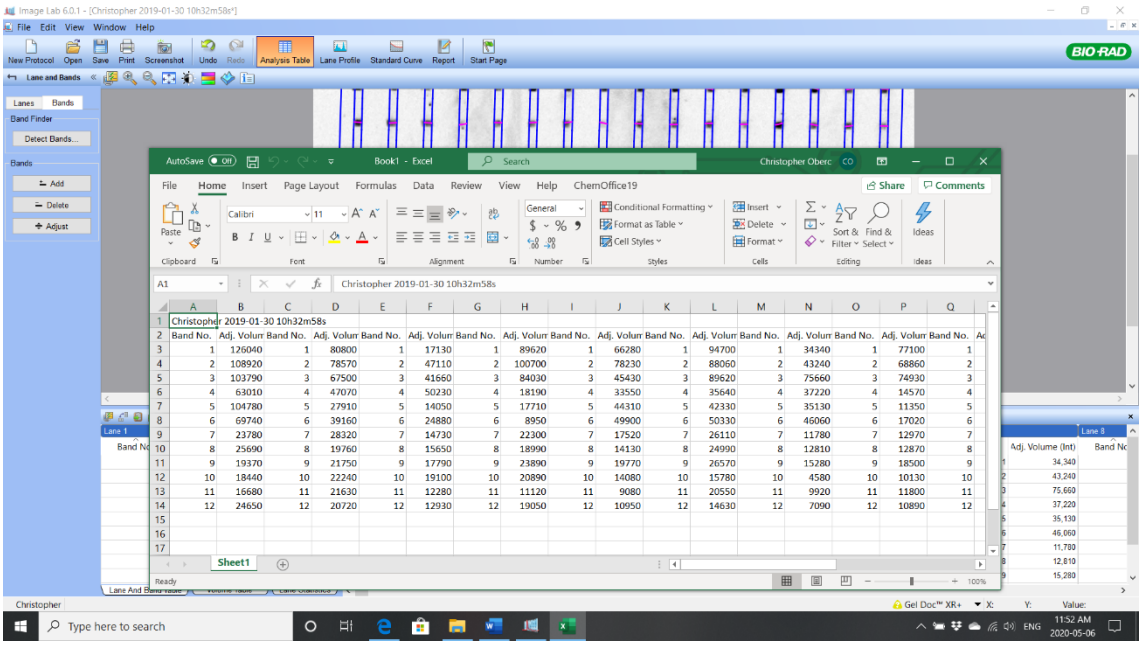
Step 12: Next, select the “Analysis Tables” on the top menu bar, and a window was opened which displayed a table of the bands. The number of values was checks in each column to make sure that it matched the number of horizontal lanes used. Any discrepancy was an indication that either a band was missing, or a false positive band was not deleted.



Step 13: On the top left on the table, the first button was opened, and in the popup menu, “Band No” and “Adj. Volume (Int)” measurements were put in the right column “Displayed”, while all other measurements remained in the left column “Not Displayed”. Click on “Ok” was clicked on to confirm these measurements in the table. The importance of having the “Band No.” present was that without this metric, the program will automatically arrange the values in each column in descending order. Having “Band No.” maintains the value order as presented in the array.



Step 14: Select the fifth button in the top left corner of the table was selected which copied the table to an Excel file.



Step 15: Once the Excel file was generated, all the “Band No.” columns were deleted since they are no longer needed (Excel does not automatically rearrange values).

Appendix B: Melting Temperature Data

The melting temperatures of the duplexed formed between the oligonucleotide probes (N1Q, N1G, N2Q, N2G) and the targets (Gin, Quin) were measured. The experimentally measured values are compared with the theoretically calculated values.

Experimental melting curves were measured using a UV spectrometer (Evo260) with a temperature controller (PCCU1). All test solutions contained 250 nM of each of the two hybridizing strands in the hybridization buffer (1× SSC, 0.1% SDS).

Duplex	**Theoretical Tm (°C) from DINAMelt**	**ΔTm of Theoretical (PM-MM) (°C)**	**Experimental Tm (°C) from UV measurement**	**ΔTm of Experimental (PM-MM) (°C)**
N1G-Gin	60.0	3.0	61.4	2.3
N1G-Quin	57.0		59.1	
N1Q-Gin	56.2	4.6	57.3	5.0
N1Q-Quin	60.8		62.3	
N2G-Gin	61.5	4.9	63.2	6.9
N2G-Quin	56.6		56.1	
N2Q-Gin	57.5	3.1	58.2	3.4
N2Q-Quin	60.6		61.6	

Theoretical and experimental Tm values of the different combinations of perfectly complementary and 1 bp mismatching oligonucleotides. DINAMelt[29] was used to determine the theoretical Tm values using 0.15 M Na^+ and the experimental Tm values were determined from the inflection points of the melting curves.

Appendix C: Raw Fluorescence Microarray Images Produced in Multiplex Nucleic Acid Tests

This appendix collected the raw image data for creating the bar graphs presented in Chapters 3 and 4 of the thesis.

Figure 7.1: The relative intensities of biotin-labelled PCR products (Asymmetric, 226 nt) of the six ginseng powder samples hybridizing to immobilized A-N1G, A-N2G, A-N1Q and A-N2Q (Method 1). The relative intensities have been normalized with Int selected in the consensus region as an internal standard (see Figure 2.3). Bar graph is on Figure 3.3.

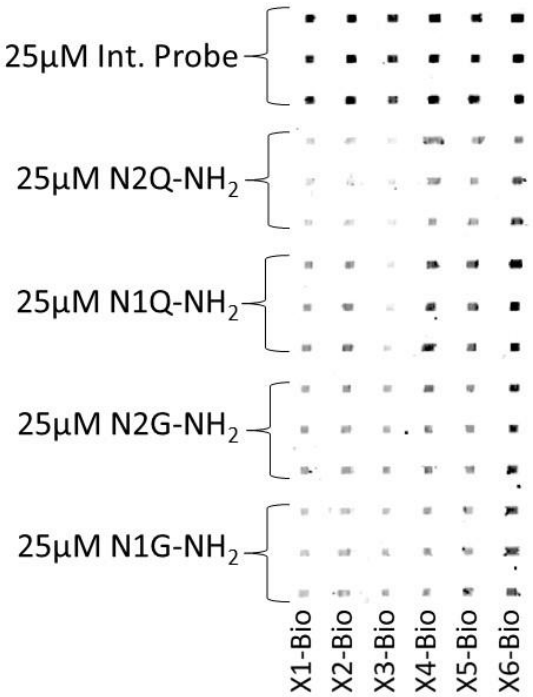


Figure 7.2: Fluorescence intensity of biotin-labeled X3 and AmG3 PCR strands (226 nt) hybridized to the immobilized probes: N2G, N1Q, tN1Q, N3Q, and LN3Q (Method 1). The relative intensities have been normalized with P8'. The two images are from the same experiment; the right image has improved contrast relative to the left image. Bar graph is on Figure 3.3.

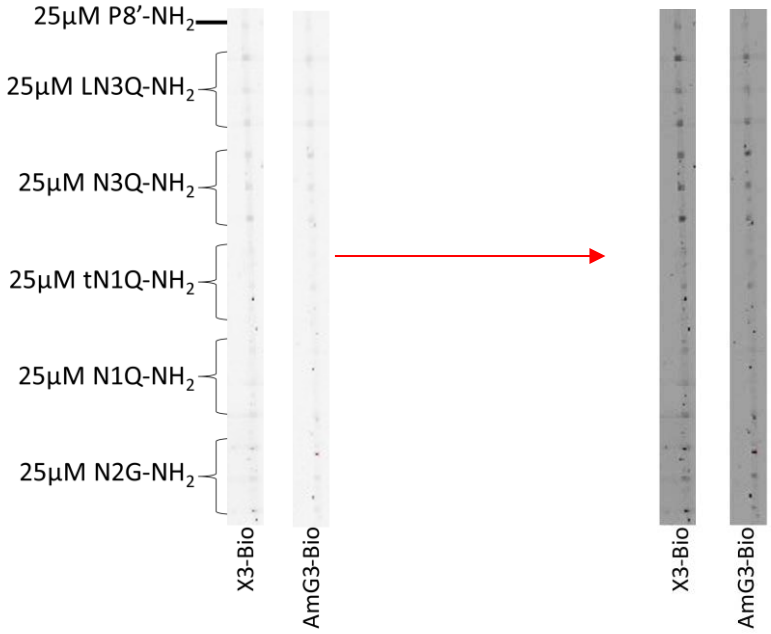


Figure 7.3: Fluorescence intensity of biotin-labeled X3 and AmG2 (126 nt, 226 nt, 694 nt) PCR strands hybridized to the immobilized probes: N2G, tN2G, tN1Q and N3Q (Method 1). This graph is normalized to factor in the differences in MW of the PCR strands. The inset shows the results in the expanded scale for the 694 nt PCR strand. The bar graph is on Figure 3.4.

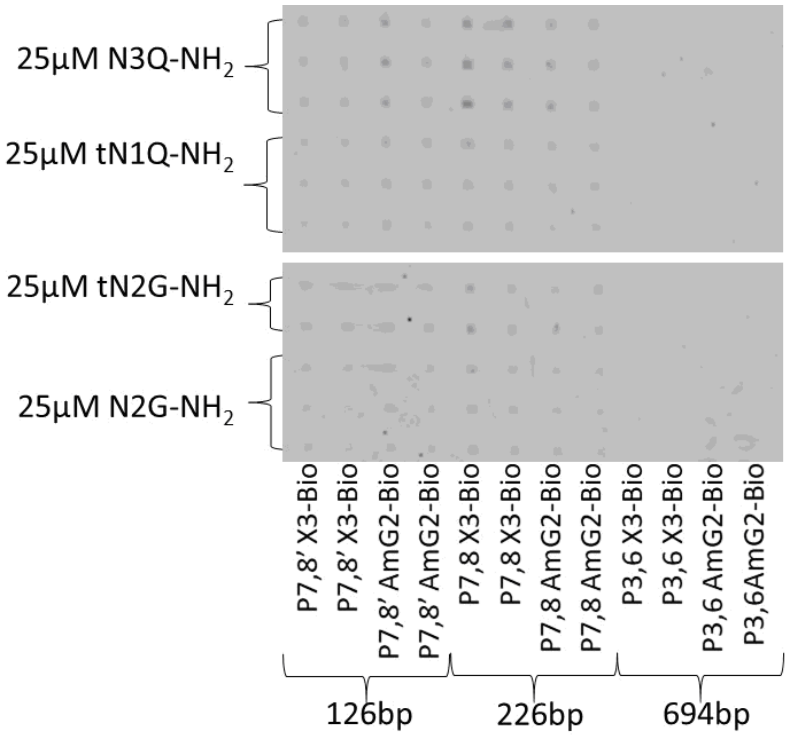


Figure 7.4: Immobilization of 226 nt AmG2 PCR strand to the solid surface using $NaCNBH_3$ and incubating it for 2, 4 and 6 hrs.: B-N2Q was used for detection. The bar graph is on Figure 3.6.

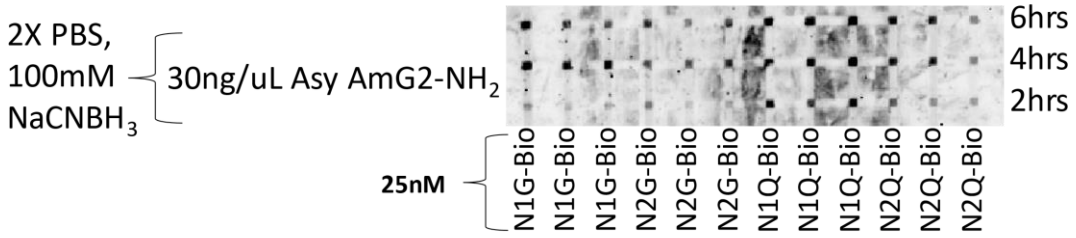


Figure 7.5: Hybridization of immobilized A-Gin and A-Quin to B-N2G and B-N1Q (Method 2), a) SF hybridization, b) dynamic hybridization (1 injection of 1.0 µL, 5 spins). 100 nM oligo was immobilized. The bar graph is on Figure 3.9. The image on the right is the right half of the image to the left with increased contrast.

bottom: Hybridizations of immobilized 226 nt X3 and AmG2 PCR strands to B-N1G, B-N2G, B-N1Q and B-N2Q detection probes (Method 2): a) SF hybridization (1 h) b) Centrifugal (dynamic) hybridization using a single 1.0 µL injection. The bar graph is on Figure 3.9b The image on the right is the right half of the image to the left with increased contrast.

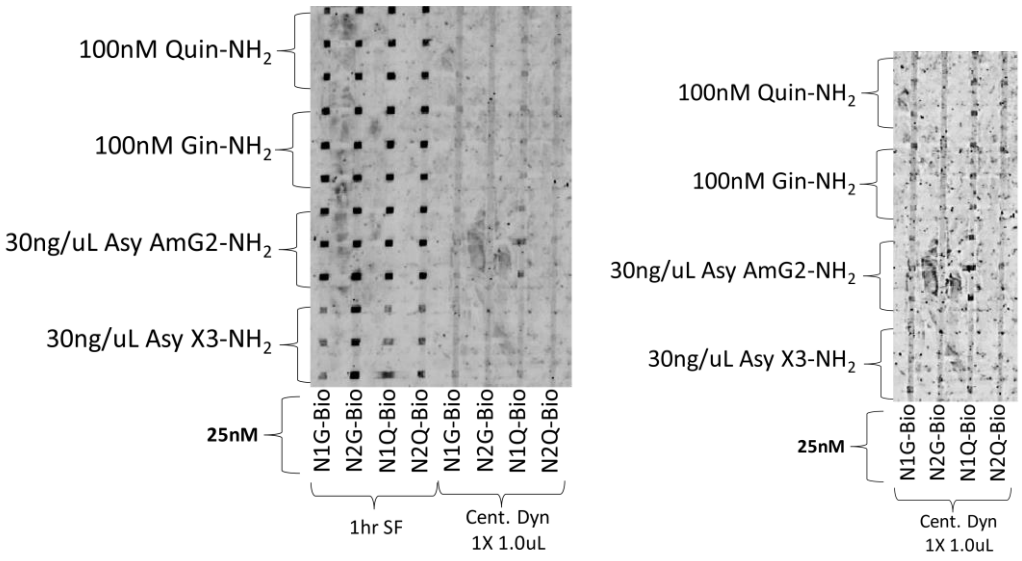


Figure 7.6: SF hybridization of immobilized 226 nt X3 and AmG2 to B-N2G, B-tN2G, B-N1Q, B-tN1Q and B-N3Q (Method 2). The hybridization was performed using the 1 h SF method. The bar graph is on Figure *3.9*a and Figure *3.10*.

(top and bottom) All hybridization was done using the 1 h SF method. X3 and X3' are the 226 nt and 126 nt PCR strands, respectively; AmG2 and AmG2' are the 226 nt and 126 nt PCR strands, respectively. The intensities were used to construct the graph in Figure *3.10* and

Figure 3.11.

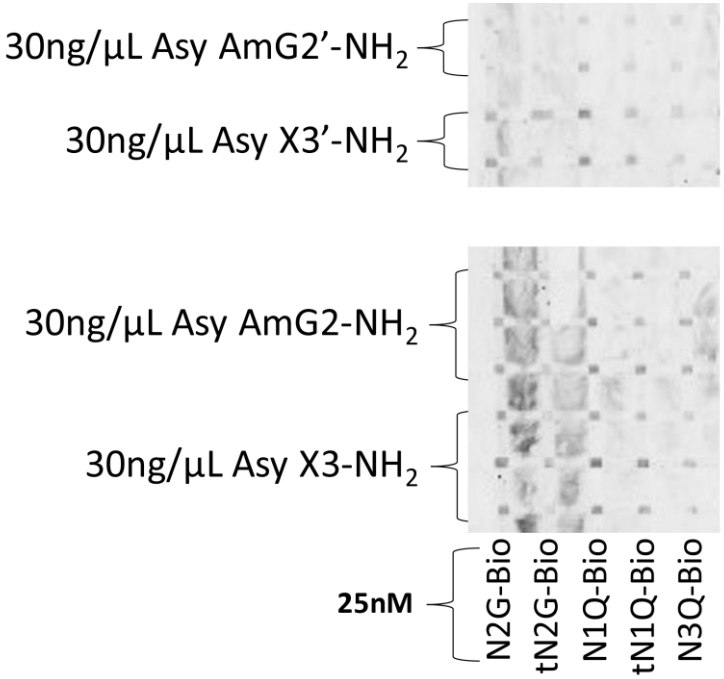


Figure 7.7: a) Three strand complexes made by hybridizing unlabeled 226 nt asymmetric X3 and AmG2 to immobilized 5'-A-P8' capture strand (Method 3). Then B-N2G, B-tN2G, B-N1Q, B-tN1Q, B-N3Q are hybridized to those captured PCR strands (SF, 1 h). Each three-strand complex was done is six replicates. The bar graph is on Figure 3.13.

Three strand complexes made by hybridizing unlabeled 226 nt X3 and AmG2 to immobilized 3'-A-P8' capture strand as the capture strand (Method 3). Then B-N2G, B-tN2G, B-N1Q, B-tN1Q, B-N3Q are hybridized to those captured PCR strands. b) Both hybridization steps were performed using SF hybridization c) The PCR strand hybridization was performed using SF hybridization and the probe hybridization was performed using dynamic hybridization. All three-strand complex were made in six replicates. The bar graph is on Figure 3.16.

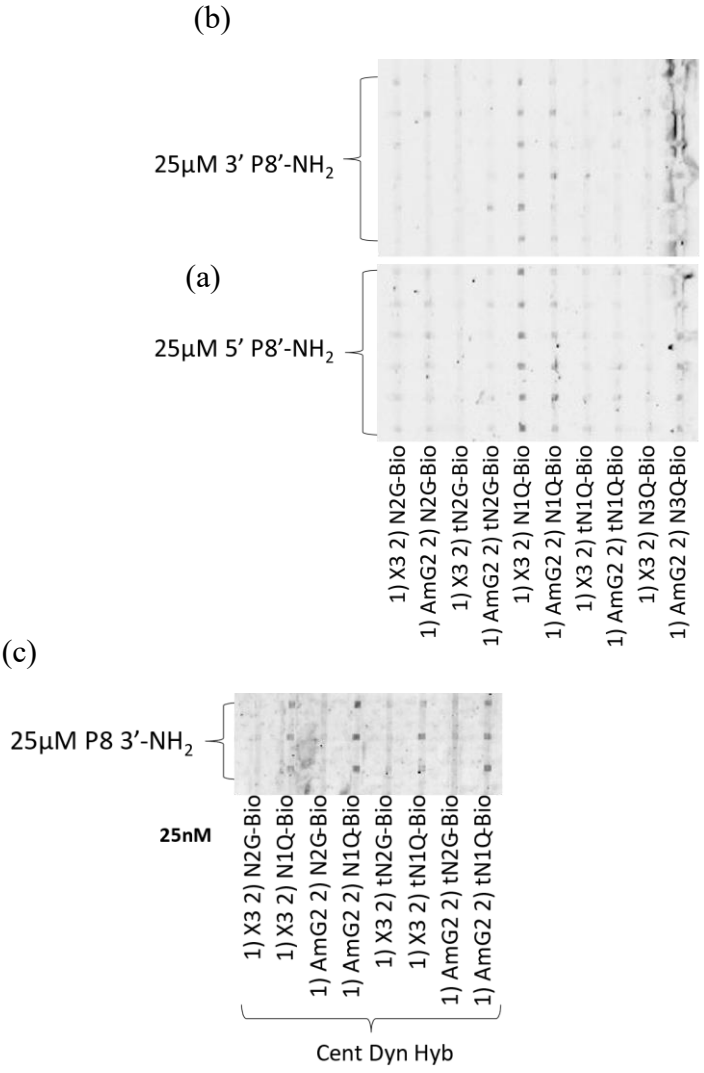


Figure 7.8: Three strand complexes made by hybridizing unlabeled X3 and AmG2 (40 ng/μL) to immobilized 5'-A-P8' capture strand (Method 3). Then B-N2G, and B-N1Q are hybridized to those captured PCR products. Each three-strand complex was done in six replicates and both hybridization steps were done using SF hybridization. The top graph was constructed using the fluorescence intensities and the bottom graph was constructed from the quotient of those same intensities over the strand's concentration in nanomolar. The bar graph is on Figure 3.14.

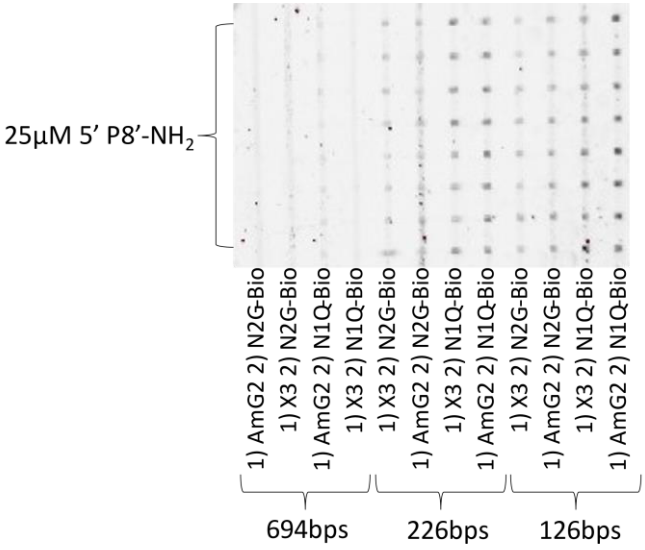
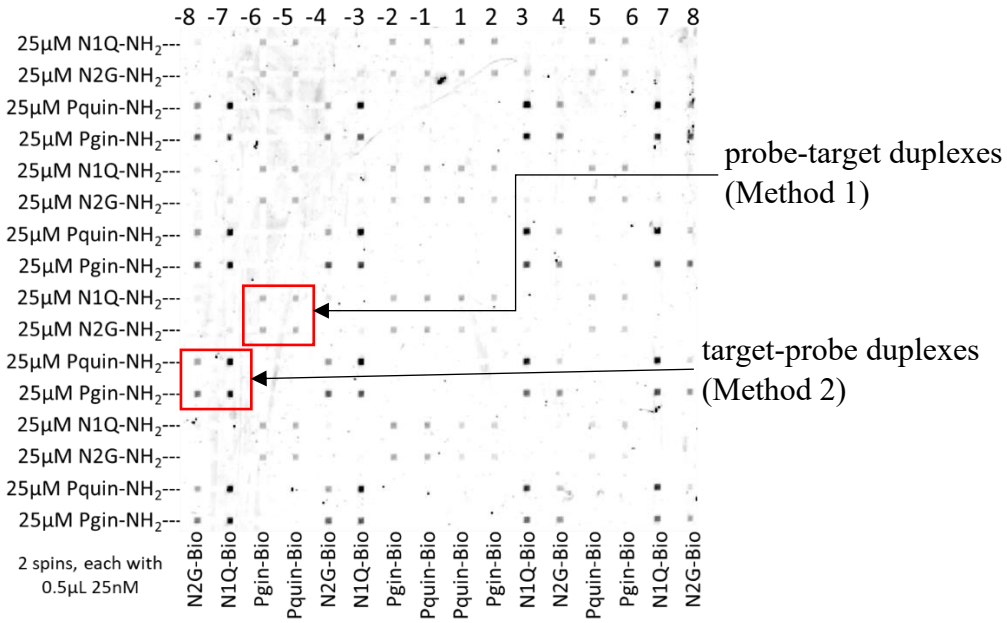


Figure 7.9: Hybridization of Gin/Quin targets to the N2G and N1Q probes using 2 injections of 0.5 µL. The channel numbering refers to the numbering used in Figure 3.17. Each hybridization was made in quadruplicates. a) hybridization of immobilized targets to biotin labeled probes (Method 2). b) Hybridization of immobilized probes to biotin labeled targets (Method 1). 25 µM oligonucleotide solutions were used for immobilization. The bar graph is on Figure 3.18.



All identical duplexes in the same column were used to calculate the average and standard deviation.

Figure 7.10: Comparison of the target-probe Method 2 using one spin, target-probe Method 2 using five spins and the probe-target Method 1 using five spins. The signals for the probe-target Method 1 using one spin were so low that they were not shown. The bar graph is on Figure 3.19.

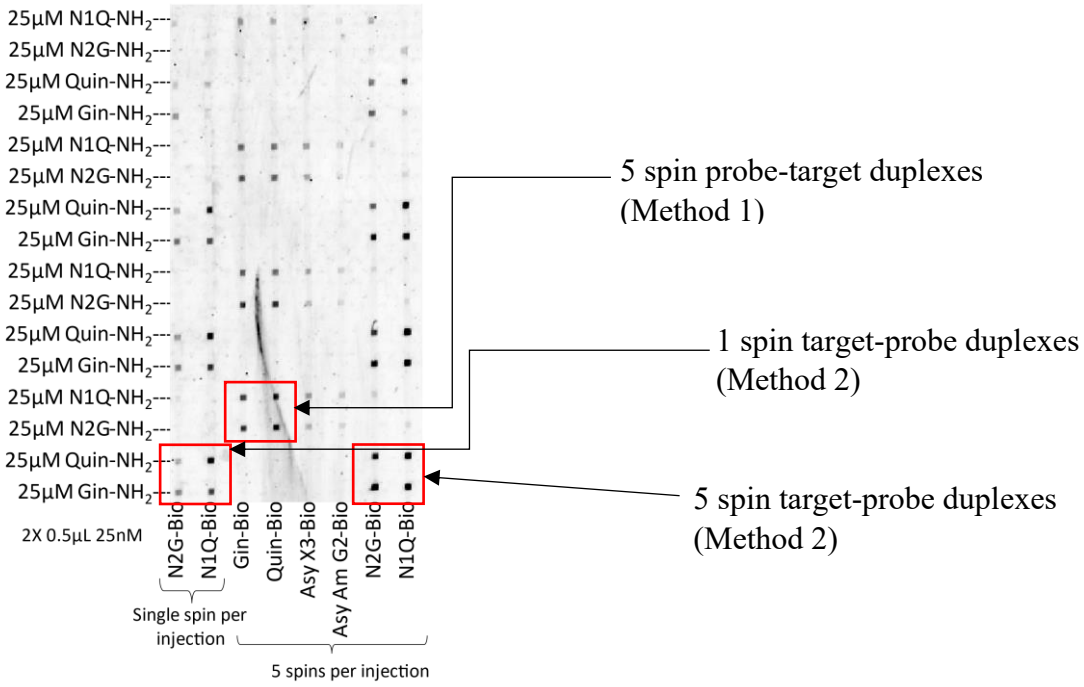


Figure 7.11: Centrifugal probe-target Method 1 hybridizing asymmetric biotin labeled 226 nt X3 and AmG2 to immobilized N2G and N1Q. The five-spin method was used with two injections of 0.5 µL each. Each injection was spun five times and the solutions was allowed to backflow after each spin. The hybridization buffer contained a) no additives, b) 2.5% formamide (v/v), c) 5.0% formamide (v/v) or d) 10% formamide (v/v). The bar graph is on Figure 3.20.

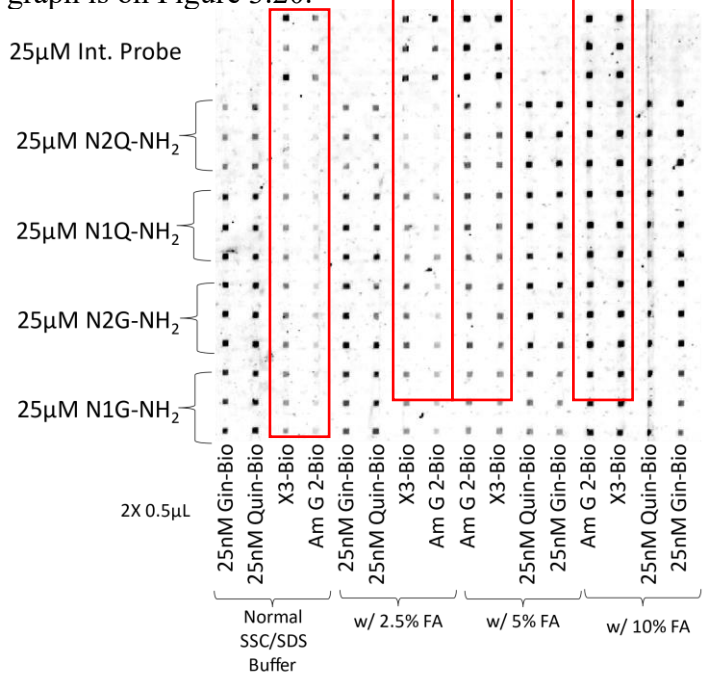


Figure 7.12: Centrifugal probe-target Method 1 hybridizing asymmetric biotin labeled 226 nt X3 and AmG2 to immobilized N2G and N1Q. The five-spin method was used with two injections of 0.5 µL each. Each injection was spun five times and the solutions was allowed to backflow after each spin. The hybridization buffer contained a) no additives, b) 2 M urea, c) 4 M urea or d) 6 M urea. Although the use of 4 M urea (c) increases differentiation by N2G, this does not produce adequate differentiation for N1Q. Both the chip image and bar graphs are shown.

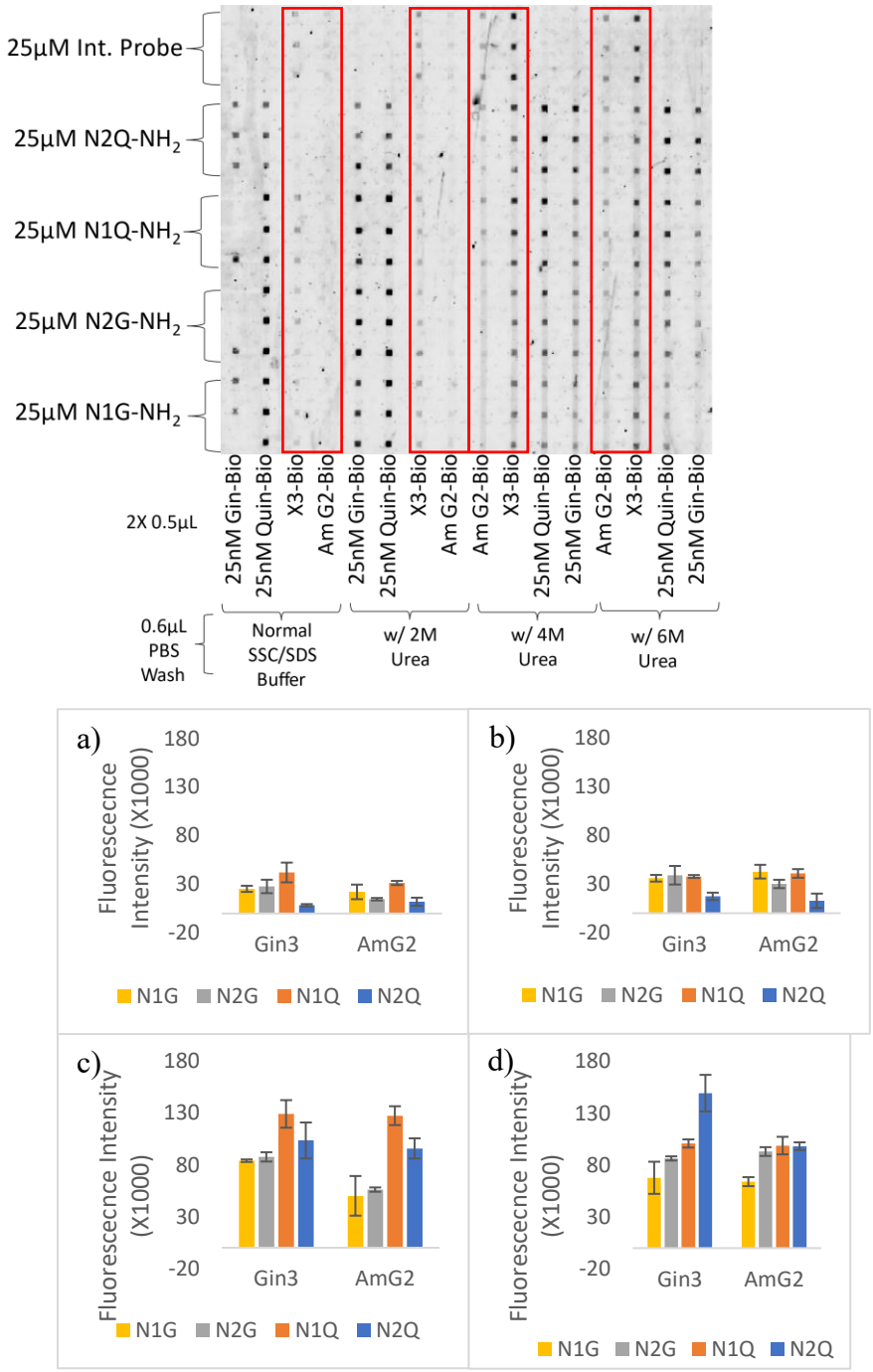


Figure 7.13: Effects of reaction time on LIDA. Three reaction times were used for the ligation reaction with Quin-NH_2 and Quin gBlock ds samples as templates using QSF. All the data bars are the averages of quadruplicates, and the error bars are the standard deviations. NEB ligase was used. The bar graph is on Figure 4.2.

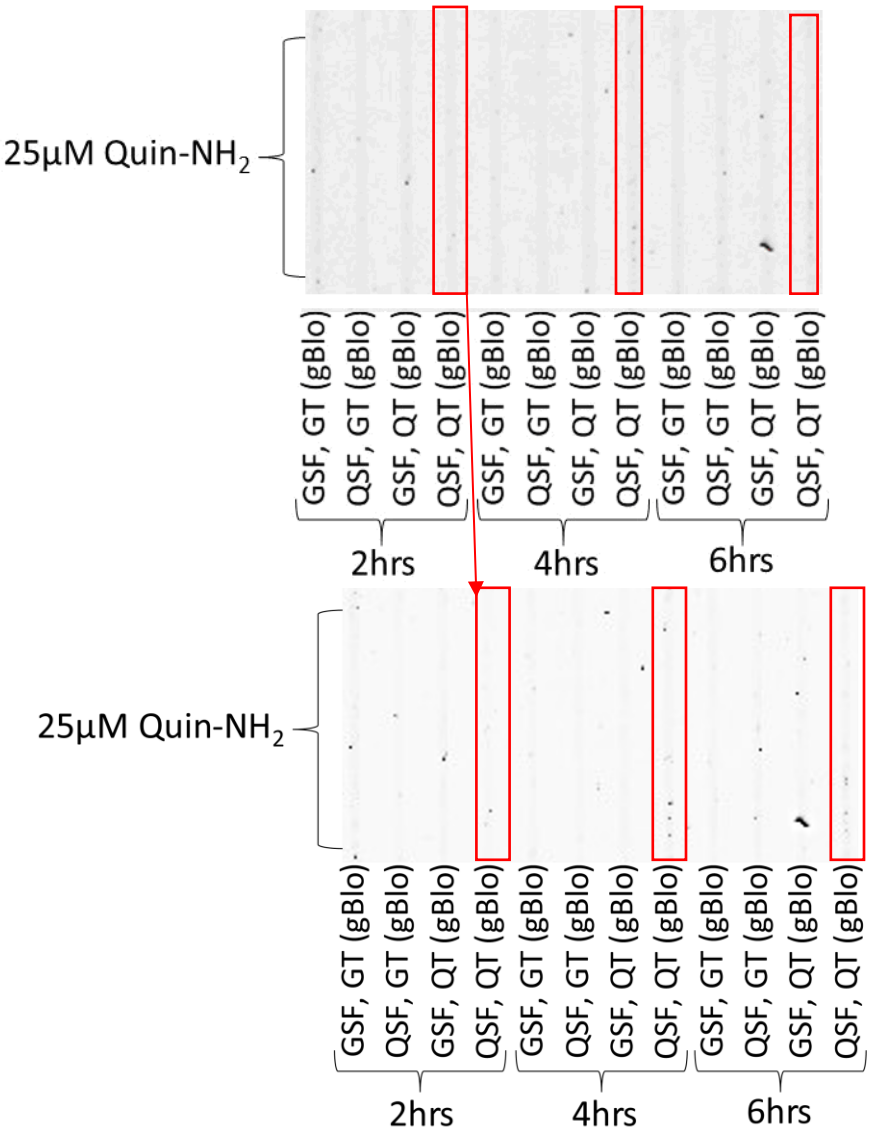


Figure 7.14: The effect of adding free-solution Gin-NH_2 or Quin-NH_2 in the ligation mixture. Gin-NH_2 or Quin-NH_2 (10 nM) were added to the ligation mixture, to be compared with controls containing no free-solution Gin-NH_2 or Quin-NH_2. The mixture was incubated for (a) 2 h, (b) 4 h. AmG2 genomic templates and 10 mM Mg^{2+} were used for all ligation reactions. The bar graph is on Figure 4.3.

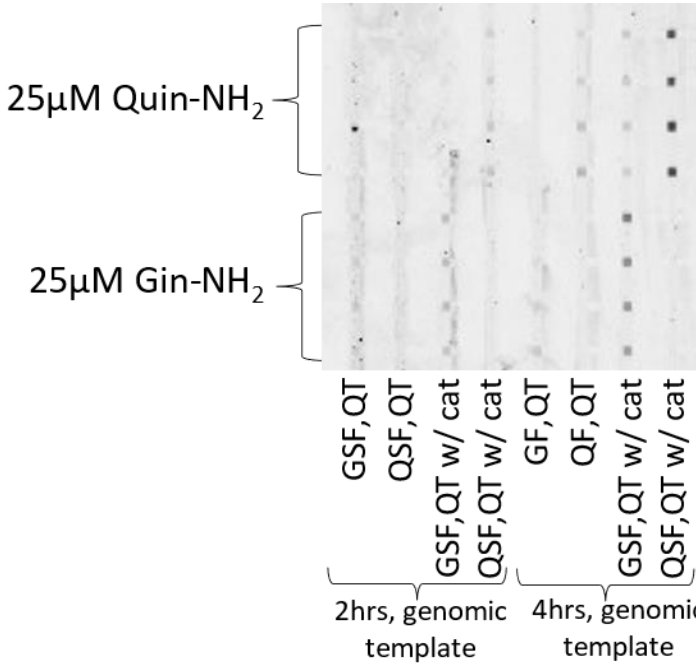


Figure 7.15: LIDA of gBlock templates. 100 CEU/μL NEB ligase was used, and incubation time was 4 h. 10 nM Gin-NH_2 or Quin-NH_2, and 25 μM of GSF/QSF added. The bar graph is on Figure 4.4.

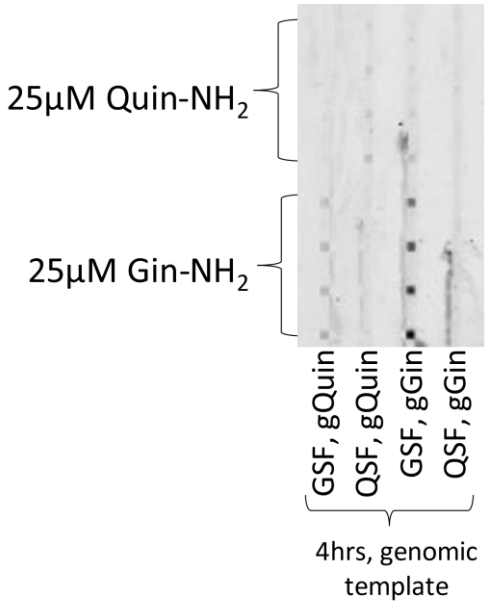


Figure 7.16: Effect of [Mg^{2+}] on LIDA. The ligation buffer already contains 10 mM Mg^{2+} and so additional Mg^{2+} was added to the buffer to result in the final concentrations of 15, 20, 25 and 30 mM. The blue bars are measured using the left vertical axis and the orange bars are measured using the right vertical axis. The bar graph is on Figure 4.5.

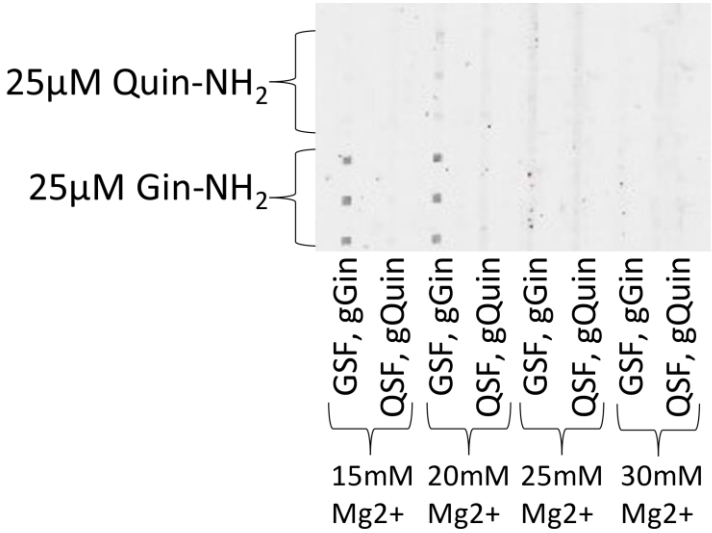


Figure 7.17: The effect of [QSF] on gQuin signals on Quin-NH_2 lanes. gQuin (1.0 pg/µL) was used as the template; Quin-NH_2 was added; 100 CEU/µL ligase was used with 4 h incubation. The last bar pertains to 25 µM GSF used with 1.0 pg/µL gGin on Gin-NH_2 lanes. The bar graph is on Figure 4.6.

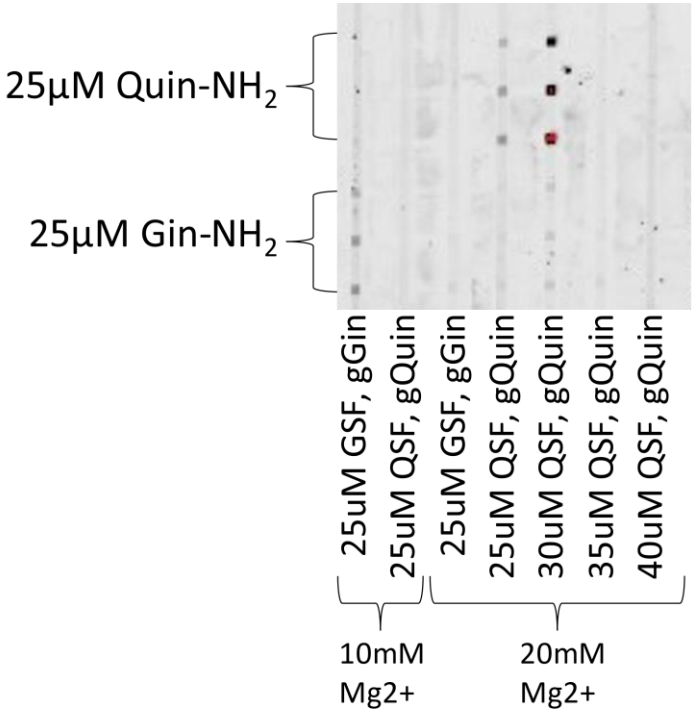
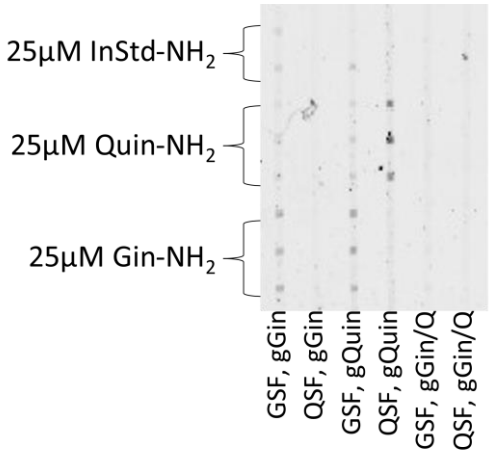


Figure 7.18: Intensity for 100% gGin, 100% gQuin, and a 50% mix of each (gGin/Quin). The LIDA mixture was composed of 1 µL 2× Oligo/buffer solution (Det, Frag (50 µM each), and 60 µM GSF or QSF, Gin-NH_2 or Quin-NH_2 (20 nM), 2× buffer, 20 mM $MgCl_2$), 0.5 µL template and 0.5 µL ligase (2 µL final volume). The ligation reaction was 4 h. a) 100 CEU/µL NEB ligase b) 1.25 U/µL TF ligase. The bar graph is on Figure 4.7.

(a)



(b)

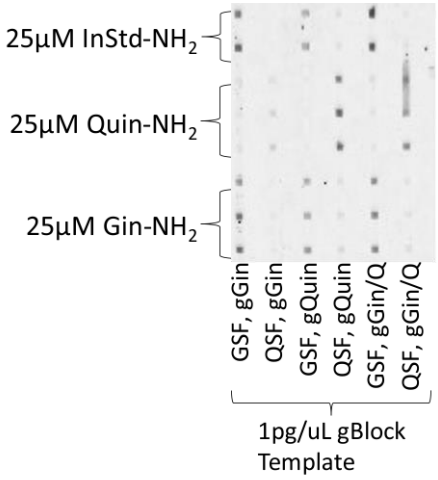


Figure 7.19: Intensity for 100% gGin, 100% gQuin, and a 50% mix of each (gGin/Quin). The LIDA mixture was composed of 1 µL 2× Oligo/buffer solution (Det, Frag (50 µM each), and 60 µM GSF or QSF, Gin-NH_2 or Quin-NH_2 (20 nM), 2× buffer, 20 mM $MgCl_2$), 0.5 µL template and 0.5 µL TF ligase (2 µL final volume) and was incubated for 2 h. The bar graph is on Figure 4.8.

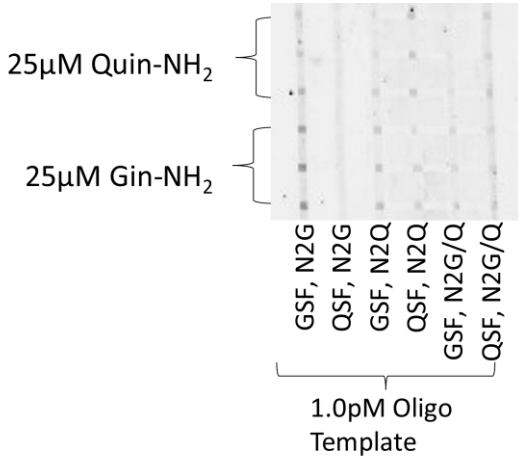


Figure 7.20: Differentiation of ginseng samples X1 to X6. Genomic templates used; 10 nM Gin-NH_2 or Quin-NH_2 added, 30 µM GSF or QSF, 25 µM Det, 25 µM Frag, 10mM [Mg^{2+}] (added), 1.25 U/µL TF ligase, 2 h incubation. The bar graph is on Figure 4.9.

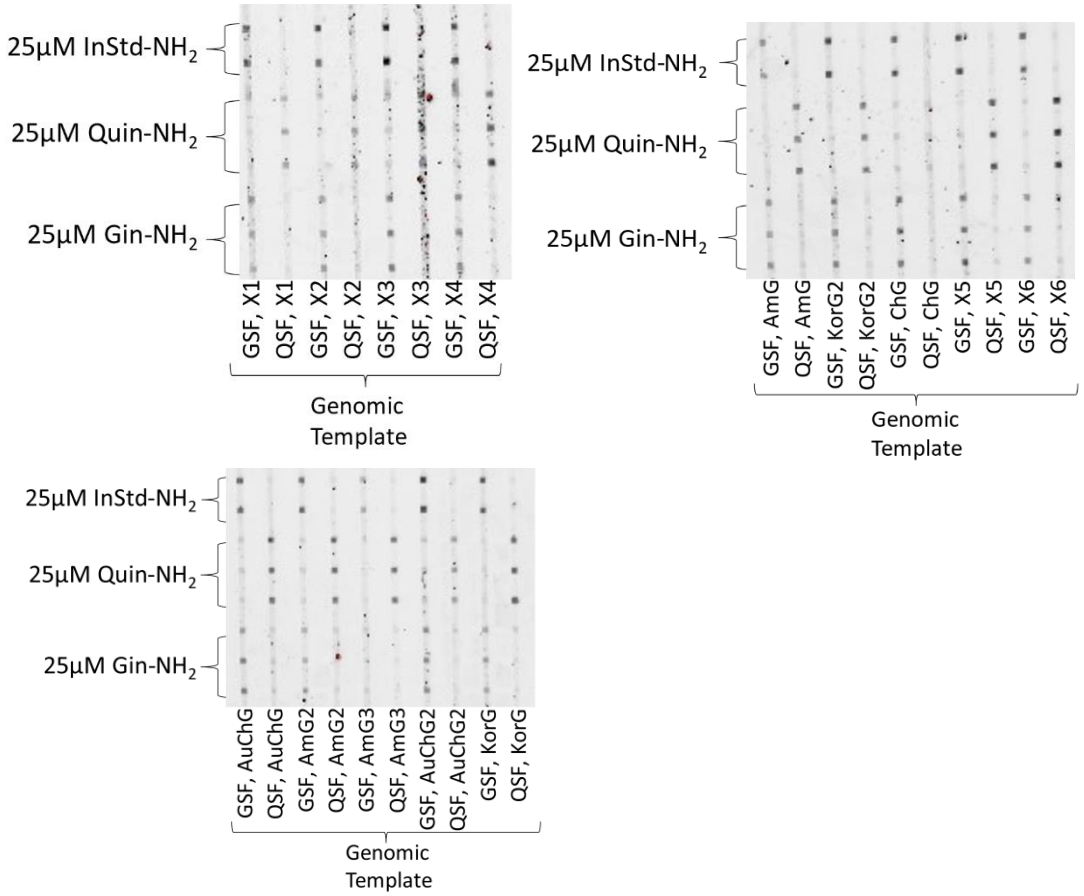
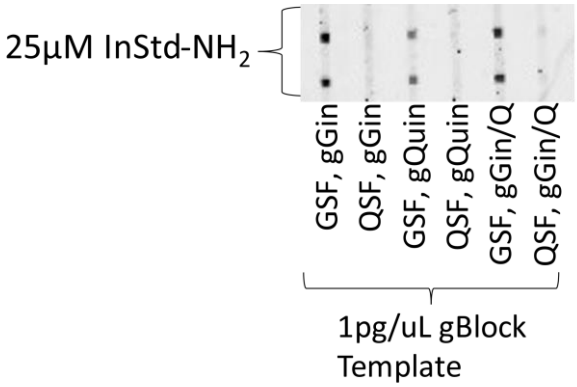


Figure 7.21: Fluorescence intensities at the immobilized fragment InStd-NH_2 used in combination with LIDA solutions containing 30 µM GSF or QSF, 25 µM Frag, 25 µM Det and 1.0 pg/µL gBlock templates. The bar graph is on Figure 4.14a.



Appendix D: Supporting Figures

Predicted Hybridization Rate Constants

Table 17: Probe hybridization rate constants predicted using http://nablab.rice.edu/nabtools/kinetics.html[36] at 22°C and 0.15 salinity. The rate constants are calculated under the assumptions that the probes are hybridizing to targets that are exact complements and that the probes and targets are both in solution phase.

Probe Name	Rate Constant (1/Ms)
N1G	8.13×10^5
N2G	2.04×10^6
N1Q	2.23×10^6
N2Q	2.03×10^6
LN3Q	4.99×10^6
tN1G	4.04×10^6
tN2G	4.68×10^6
tN1Q	4.37×10^6
tN2Q	4.61×10^6
N3Q	1.44×10^7

Centrifugal Flow Rate

Table 18: The amounts of time required for the fluid to flow from the inlet reservoir to the outlet reservoir for each of the 8 channels on one side of the chip when the chip holder was spun at 600 rpm. Each of the four directions: A, B, C, D (second column) corresponds to the 4 directions shown in Figure 2.6a. It is found that direction C is optimal.

Flow Conditions		Flow Completion Time (seconds)							
RPM	**Direction**	**Channel 1**	**Channel 2**	**Channel 3**	**Channel 4**	**Channel 5**	**Channel 6**	**Channel 7**	**Channel 8**
600	A (30°)	40	40	35	35	17	17	17	20
600	A (30°)	30	25	20	20	15	15	15	15
600	A (30°)	25	20	30	37	19	19	15	15
	Average	31.66667	28.33333	28.33333	30.66667	17	17	15.66667	16.66667
	Stand Dev	6.236096	8.498366	6.236096	7.586538	1.632993	1.632993	0.942809	2.357023
600	C (40°)	40	25	25	25	17	17	20	20
600	C (40°)	27	27	30	30	12	12	22	22
600	C (40°)	25	25	25	25	15	15	10	12
	Average	30.66667	25.66667	26.66667	26.66667	14.66667	14.66667	17.33333	18
	Stand Dev	6.649979	0.942809	2.357023	2.357023	2.054805	2.054805	5.249339	4.320494
600	D (50°)	45	36	30	20	20	16	15	10
600	D (50°)	50	40	30	25	17	15	15	15
600	D (50°)	55	40	35	15	12	12	10	12
	Average	50	38.66667	31.66667	20	16.33333	14.33333	13.33333	12.33333
	Stand Dev	4.082483	1.885618	2.357023	4.082483	3.299832	1.699673	2.357023	2.054805
600	B (60°)	164	158	54	45	28	26	25	22
600	B (60°)	180	160	60	35	45	23	24	22
600	B (60°)	135	150	50	38	46	32	32	25
	Average	159.6667	156	54.66667	39.33333	39.66667	27	27	23
	Stand Dev	18.62495	4.320494	4.109609	4.189935	8.259674	3.741657	3.559026	1.414214

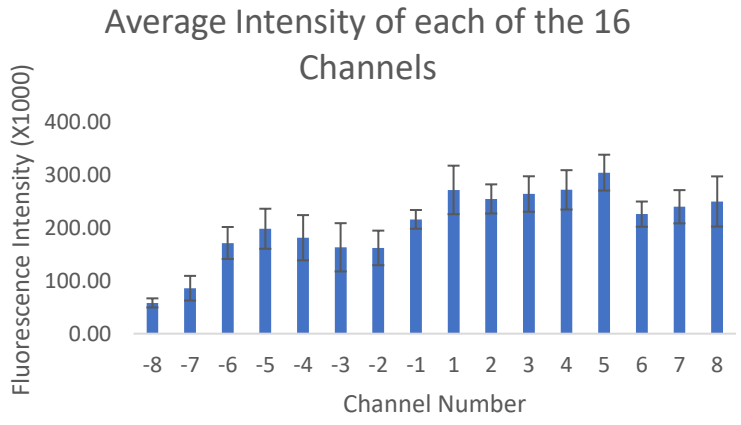


Figure 7.22: Comparison between using a single injection of 1.0 µL (negative channel numbers) and 2 injection of 0.5 µL each (positive channel numbers). Each bar represents 12 replicates of hybridization of Quin-NH_2 to N1Q-Bio lanes.

Table 19: t-Test of Figure 3.18b Probe-Target (Method 1) Results

N2G-Gin

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	64829.35	62179.77
Variance	24918595	2.66E+08
Observations	3	3
Hypothesized Mean Difference	0	
df	2	
t Stat	0.269283	
P(T<=t) one-tail	0.406474	
t Critical one-tail	2.919986	
P(T<=t) two-tail	0.812949	
t Critical two-tail	4.302653	

N2G-Quin

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	65761.82	74402.97
Variance	25036186	1.44E+08
Observations	3	3
Hypothesized Mean Difference	0	
df	2	
t Stat	-1.1511	
P(T<=t) one-tail	0.166556	
t Critical one-tail	2.919986	
P(T<=t) two-tail	0.333112	
t Critical two-tail	4.302653	

N1Q-Gin

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	69889.45	64925.35
Variance	14279263	1.83E+08
Observations	3	3
Hypothesized Mean Difference	0	
df	2	
t Stat	0.61273	
P(T<=t) one-tail	0.301222	
t Critical one-tail	2.919986	
P(T<=t) two-tail	0.602445	
t Critical two-tail	4.302653	

N1Q-Quin

t-Test: Two-Sample Assuming Unequal Variances

	Variable 1	*Variable 2*
Mean	68451.51	78034.72
Variance	7863042	41818274
Observations	3	3
Hypothesized Mean Difference	0	
df	2	
t Stat	-2.35492	
P(T<=t) one-tail	0.04993	
t Critical one-tail	2.919986	
P(T<=t) two-tail	0.099859	
t Critical two-tail	4.302653	

Figure 3.18b t-test comparison between each of the 4 different Probe-Target duplexes (N2G-Gin, N2G-Quin, N1Q-Gin and N1Q-Quin) (Method 1) conducted dynamically in channels -6,-5 and in channels -2,-1. In all cases, there are no statistically significant differences, i.e. p value > 0.05 for the two-tailed values. The bar graph is shown in (Figure 3.18b).

The effects of post-hybridization wash in Method 1

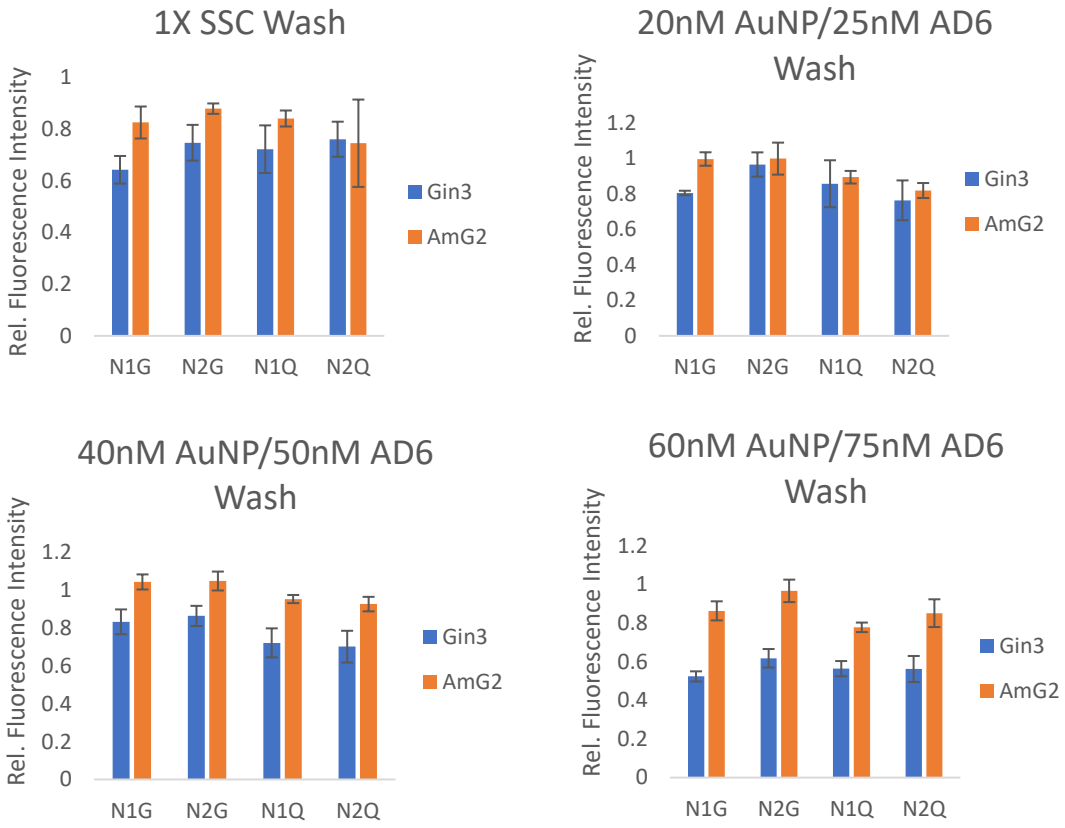


Figure 7.23: Comparison between different wash solution in Method 1 SF hybridization Different conditions are used in post-hybridization wash (1× SSC, 1× SSC with 20 nM AuNP and 25 nM AD6, 1× SSC with 40 nM AuNP and 50 nM AD6 and 1× SSC with 60 nM AuNP and 75 nM AD6). AD6 is a DNA oligonucleotide with an irrelevant sequence.

Effect of NEB Ligase Concentration in LIDA

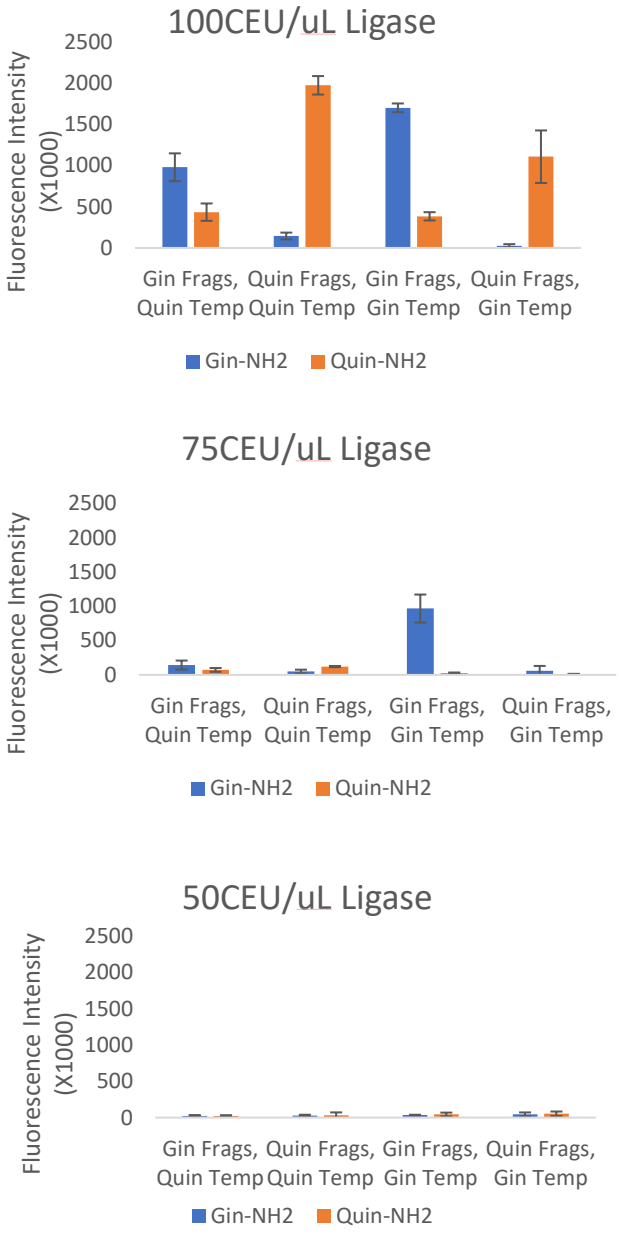


Figure 7.24: Signal intensities obtained in LIDA using different concentration of NEB ligase. The use of 100 CEU/µL ligase produced the highest signals.

Effect of Ligase Dilution Methods in LIDA

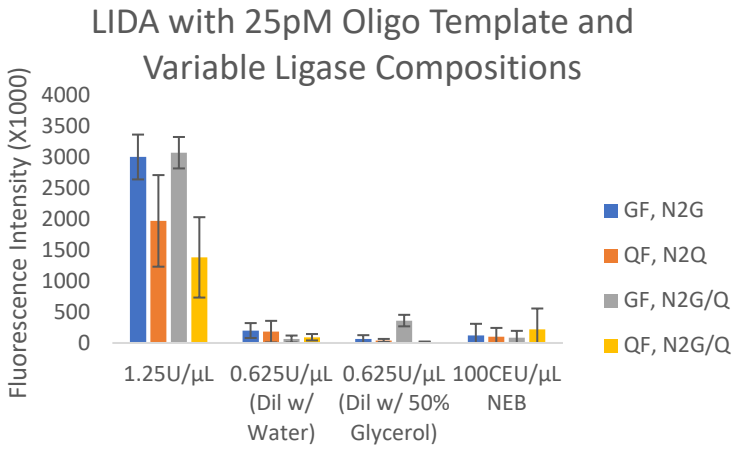


Figure 7.25: Signal intensities obtained in LIDA using undiluted TF ligase (1.25 U/µL), TF ligase diluted with either water or 50% glycerol or undiluted NEB ligase (100 CEU/µL).

Further Optimization of LIDA based on TF ligase

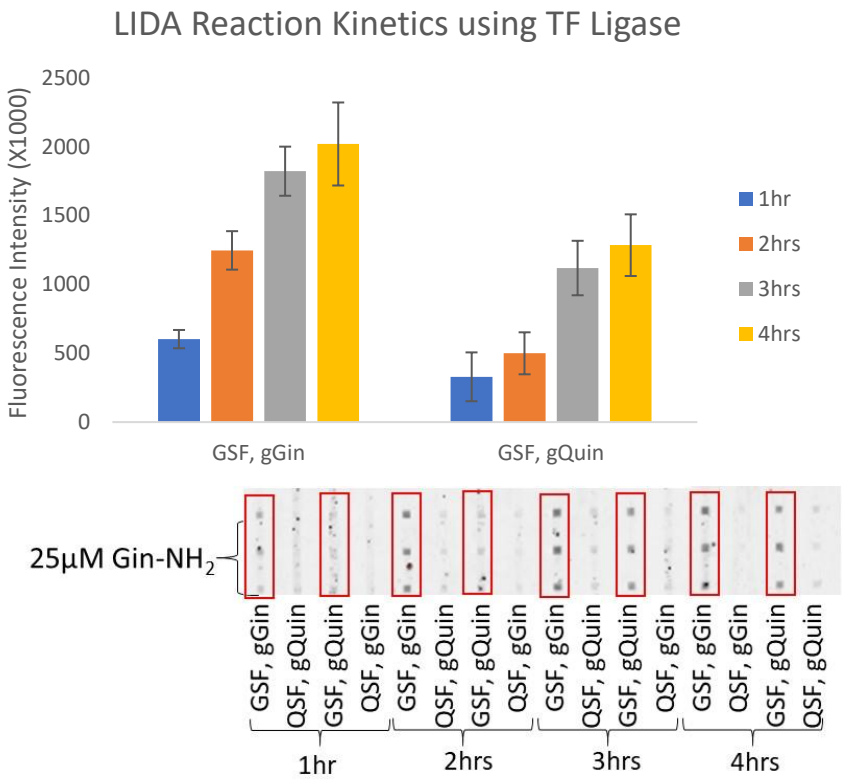


Figure 7.26: Signal intensities obtained in LIDA from different reaction times for GSF-containing lanes with either gGin or gQuin template. The ligation reactions contain Gin-NH_2 (10 nM), GSF (30 μM), Ab (25 μM), Frag (25 μM), $MgCl_2$ (20 mM) and 1.25 U/μL TF ligase.

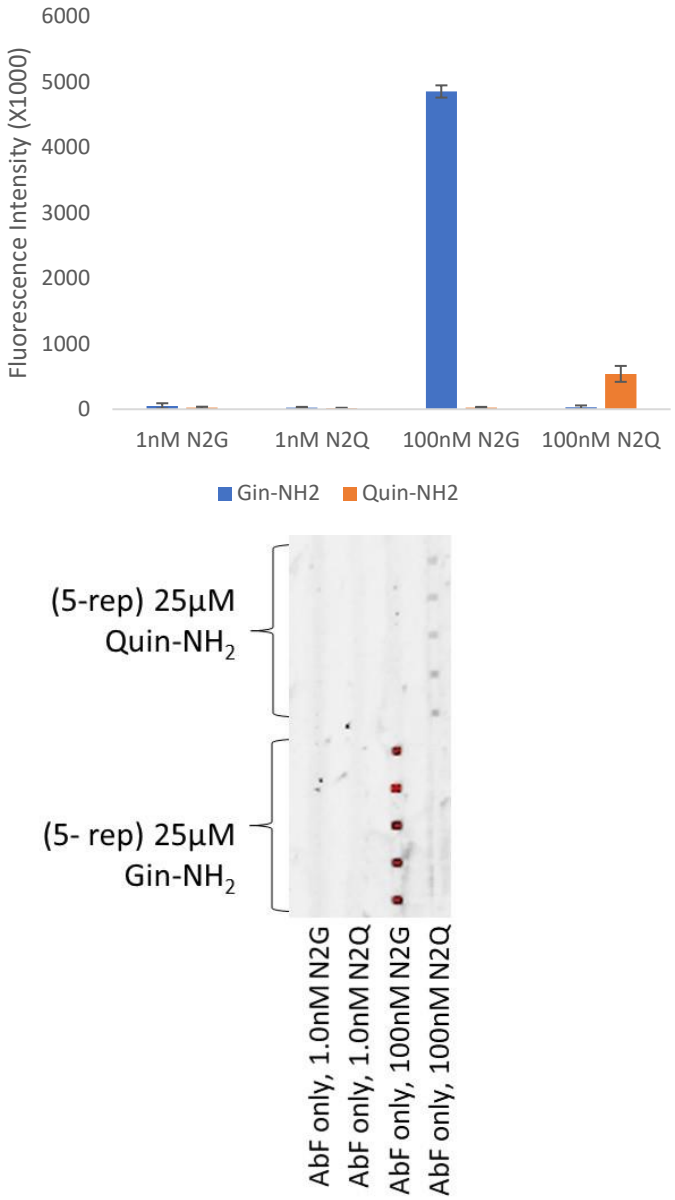


Figure 7.27: Effect on using only Cycle 1 in the LIDA reaction. Since without Cycle 2, there is no chain reaction to increase the template amounts, and so 100 nM, but not 1 nM, oligonucleotide template (A-N2G, A-N2Q) is required for notable fluoresence signal. The differentiation ratios of Gin to Quin obtained for N2G and N2Q are 160 and 22, respectively.

Appendix E: Sanger Sequencing Results of PCR products

Extractions of genomic DNA from powder sample X3 and root sample AmG2 were performed; PCR products were produced. The purified products were sent for Sanger sequencing and alignments were achieved using the *P. ginseng* sequence previously published[99] as the reference (accession number JN596111).

While the 694 nt strand is shown in black, the 226 nt strand is shown in red. The three SNP sites N1, N2, N3 are shown in blue.

X3 (sense strand X_3)

```
Gin_3-P3_G03.ab1=Sequencing, P.ginseng= Macau Reference

Gin_3-P3_G03.ab1
------------------------------------------------------------
P.ginseng
CGTTTAGATTAGATGAAAACGAACAAGTGAACTACGATGCAGTTACAACCGCTGTGAAGA


Gin_3-P3_G03.ab1
--NNNNNNNNNNNNNNNNNNNNNNNNNNNNNANCCCGATGGNCACT-GGCCANCTGAANTG
P.ginseng
AAGCTCTTCGATTGAACCGGGCAATTCAAGCACACGATGGTCACTGGCCAGCTGAAAATG
                                * ****** **** * *      ** **

Gin_3-P3_G03.ab1
CAGGCTCTTTACTTTNNACNCCTCCCCTTGTAAGTTATTAGTAAGCATTAATAATGCANC
P.ginseng
CAGGCTCTTTACTTTATACACCTCCCCTTGTAAGTTATTAGTAATCATTAATAATGCAGC
***************  ** ************************ ************* *

Gin_3-P3_G03.ab1
ACGTACGTATTGCAATGAAATCTTAATTAATTAATAATTGCATGCGCGTGTCTCTCTCTA
P.ginseng
ACGTACGTATTGCAGTGAAATCTTAATTAATTAATAATTGCATGCACGTGTCTCTCTCTA
************** ****************************** **************

Gin_3-P3_G03.ab1
TATATGTAATTATATAAGTTAATGTTTCAATATGCAGATCATTGCCCTATATATCAGCGG
P.ginseng
TATATGTAATTATATAAGTTAATGTTTCAATATGCAGATCATTGCCCTATATATCAGCGG

************************************************************

Gin_3-P3_G03.ab1
AACGATTGACACTATTCTGACAAAACAACACAAGAAGGAACTGATTCGCTTCGTTTACAA
P.ginseng
AACGATTGACACTATTCTGACAAAACAACACAAGAAGGAACTGATTCGCTTCGTTTACAA

************************************************************

Gin_3-P3_G03.ab1
CCATCAAGTAATACCAAATTAATACTCCATTCGTTCCATAATATTTGATGATTATTGAAT
P.ginseng
CCATCAAGTAATACCAAATTAATACTCCATTCGTTCCATAATATTTGATGATTATTGAAT

************************************************************
```

```
Gin_3-P3_G03.ab1
TTTTAATGTCAACTTAAGAGATTTTTAATTTCAAAATTAACATAGATTTTCTTTAAATAT
P.ginseng
TTTTAATGTCAACTTAAGAGATTTTTAATTTCAAAATTAACATAGATTTTCTTTAAATAT

************************************************************

Gin_3-P3_G03.ab1
ATACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTTTCATCTAAATTTT
P.ginseng
ATACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTTTCATCTAAATTTT

************************************************************

Gin_3-P3_G03.ab1
GAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTGGAACGGAGGGAGTA
P.ginseng
GAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTGGAACGGAGGGAGTA

************************************************************

Gin_3-P3_G03.ab1
ACAAACATGATTGAACATTGAGGTTAAAAAAAATAAATAAAATTTTGTGCTGTCCCTGCA
P.ginseng
ACAAACATGATTGAACATTGAGGTTAAAAAAAATAAATAAAATTTTGTGCTGTCCCTGCA

************************************************************

Gin_3-P3_G03.ab1
TGCTAATTAATTAAAAGGTGTTAACTGTTATGTATGTANAATG------------------
P.ginseng
TGCTAATTAATTAAAAGGTGTTAACTGTTATGTATGTAGAATGAGGATGGTGGATGGGGA
************************************ ****

Gin_3-P3_G03.ab1
-----------------------------
P.ginseng
TCCTATATTGAGGGGCACAGCACGATGAT
```

AmG2 (sense strand Query)

```
Query=Sequencing, Sbjt=Macau Reference


Query_001         1 --------------------------------------------------      0

Sbjct_001         1 TACTCCATTCGTTCCATAATATTTGATCGTTATTGAATTTTTAATGTCAA     50

Query_ 001         1 -------------------NNNNNNNNNNNNNNGGCCACNNNNNGAAGAN     31
                                                      ||       ||
Sbjct _001        51 CTTAAGAGATTTTTAATTT-----------------CA-------AA---     73

Query_ 001        32 NCCCTNCNNATCCACATCGANCCCCNNCNANNCNATACAGTGNTAATTCT     81
                          ||..||||.|| ...|   . |   . |||||||| ||||||..
Sbjct _001        74 ---------ATTAACATAGATTTTCTTTAAATATATACAGTGATAATTAA    114

Query_ 001        82 NGATTGTAACTATCTAAAAAAAATTTTCTTCTCATCTAAATTTTGAATTT    131
                      .|||||||||||||||||||..|..|||||||||||||||||||||
Sbjct _001       115 ATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTT    164

Query_ 001       132 GGACTGGTTTTAAATTGATTTTCACAACTCATATAGTTTANAACGGAGGG    181
                      |.|...|||||||||||||||.||.|||||||.|||| |||||||||
Sbjct _001       165 GAAAGTGTTTTAAATTGATTTTCAAAAGTCATATAATTTAGAACGGAGGG    214
```

```
Query_ 001       182 AGNAACAAACAAAAACNANCATAA    205
                     || ||||||||.
Sbjct _001       215 AGTAACAAACAT------------    226
```

Appendix F: Illumina NGS Supporting Data

Next generation sequencing (NGS) by Illumina (MiSeq) provided results of the six ginseng powder samples X1-6 (top), seven ginseng root samples and one ginseng tea granular sample (bottom). BWA and Java were used to analyze the some of the 28 million sequences in each of the 14 samples.

The Gin and Quin abundances were determined based on the proportion of sequences that contain a “C” and “T” at the N2 SNP site which correspond to *P. ginseng* and *P. quinquefolius,* respectively. Preliminary considerations are on *P. ginseng:* "T, C, G" and *P. quinquefolius:* "C, T, A" for the N1, N2, N3 sites, respectively.

Table 20: Abundances of X1-6 at the N2 SNP site.

X1_R1_001.fastq			X2_R1_001.fastq			X3_R1_001.fastq			X4_R1_001.fastq			X5_R1_001.fastq			X6_R1_001.fastq		
total: 25975620			total: 23578139			total: 27673377			total: 24825544			total: 26513922			total: 22963062		
T	15812567	0.61	T	14893577	0.63	C	13565445	0.49	T	16229110	0.65	T	14871880	0.56	T	14852312	0.65
C	7800421	0.30	C	6246464	0.26	T	11762526	0.43	C	6911347	0.28	C	8099565	0.31	C	6486697	0.28

Table 21:Abundances of X1-6 at the N2/N3 SNP sites.

X1_R1_001.fastq		X2_R1_001.fastq		X3_R1_001.fastq		X4_R1_001.fastq		X5_R1_001.fastq		X6_R1_001.fastq	
total: 25975620		total: 23578139		total: 27673377		total: 24825544		total: 26513922		total: 22963062	
TA	13941548	TA	12798356	CG	11511756	TA	14604229	TA	12438868	TA	13432610
CG	6171627	CG	4734791	TA	10061510	CG	5625344	CG	6082378	CG	5428715

Table 22: Abundances of X1-6 at the N1/N2/N3 SNP sites.

X1_R1_001.fastq			X2_R1_001.fastq			X3_R1_001.fastq			X4_R1_001.fastq			X5_R1_001.fastq			X6_R1_001.fastq		
total: 25975620			total: 23578139			total: 27673377			total: 24825544			total: 26513922			total: 22963062		
CTA	13066743	0.5	CTA	11672968	0.5	CTA	9186072	0.33	CTA	13875931	0.56	CTA	11244157	0.42	CTA	12894605	0.56
CCG	4724711	0.18	CCG	3036746	0.13	TCG	5524591	0.2	CCG	4744069	0.19	CCG	3650256	0.14	CCG	5015206	0.22
TCG	1196020	0.05	TCG	1476949	0.06	CCG	5517875	0.2	TCG	663042	0.03	TCG	2065386	0.08	CCA	556487	0.02
CTG	727615	0.03	CTG	524195	0.02	CTG	612035	0.02	CCA	648330	0.03	CTG	730742	0.03	CTG	464501	0.02
CCA	687631	0.03	CCA	505642	0.02	TCA	556048	0.02	CTG	551121	0.02	CCA	661944	0.02	TCG	226677	0.01

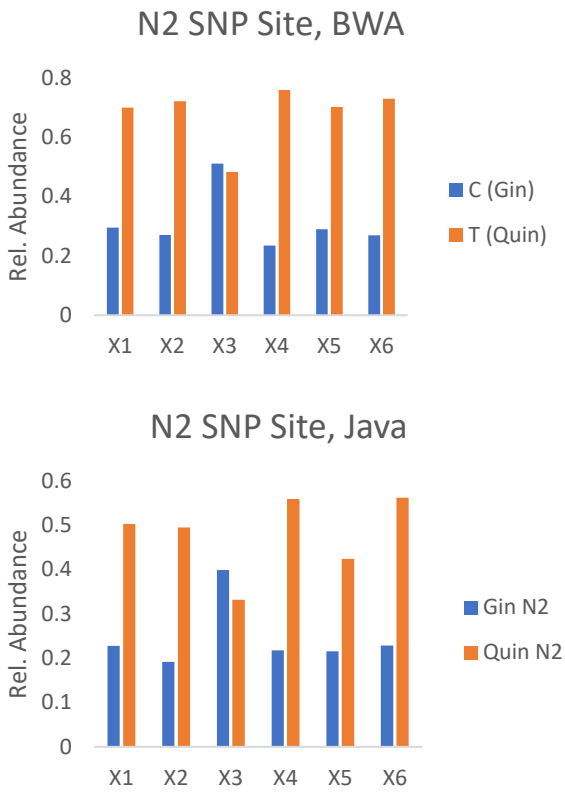


Figure 7.28: Comparison between using BWA and the Java program in analyzing the abundances in X1-6 samples at the N2 SNP site.

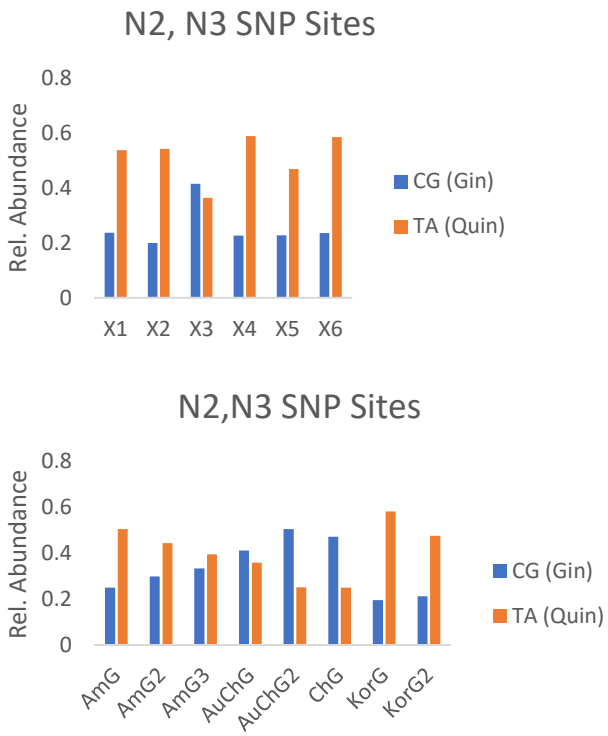


Figure 7.29: NGS results of the N2 and N3 SNP site using the Java program where *P. ginseng* contains CG and *P. quinquefolius* contains TA. All intensities have been normalized to the total number of sequences in each sample.

Table 23: Excerpt of the most abundant sequences (>1%) out of the 27,673,377 sequences of X3 using Illumina NGS. The locations of N1 (T or C), N2 (C or T) and N3 (G or A) for *P. ginseng* (TCG) and *P. quinquefolius* (CTA), respectively, are in red. The insertion observed in the CCG variants is highlighted in green. The numbers on the right represent the abundances of the sequences and the single letters (G, Q and H) represent the how we classify the sequence to be: *P. ginseng* (G), *P. quinquefolius* (Q) or CCG variant (H) based on the SNP sites.

TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTTTTAAATTGATTTTCAAAAGTCATATAATTTAGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAG	1705926	Q
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTTTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTGGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAG	1274196	G
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAATCAACAAACAATGATCATATAATTTGGAACGGAGGGAGTAACAAACATCTGTCTCTTAT	854963	H
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTTTTAAATTGATTTTCAAAAGTCATATAATTTAGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAGC	747282	Q
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTAGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAGC	683627	Q
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTTTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTGGAACGGAGGGAGTAACAAACATAACGGAGGGAGTAACAAACATAAC	547143	G
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTCTTTAAATTGATTTTCAAAAATCAACAAACAATAATTATATAATTTGGAACGGAGGGAGTAACAAACATCTGTCTCTTA	498362	H
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAATCAACAAACAATGATCATATAATTTGGAACGGAGGGAGTAACAAACATAACGGAGGGGAG	493879	H
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTTTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTGGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAGC	372853	G
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTAGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAG	344053	Q
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTTTTAAATTGATTTTCAAAAGTCATATAATTTAGAACGGAGGGAGTAACAAACATAACGGAGGGAGTAACAAACATCTG	299336	Q
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTCTCATCTAAATTTTGAATTTGAAAGTGTCTTTAAATTGATTTTCAAAAATCAACAAACAATAATTATATAATTTGGAACGGAGGGAGTAACAAACATAACGGAGGGGA	273597	H
TACAGTGATAATTAAATATTGTAACTATCTAAAAAAAAAGTATTTTTCATCTAAATTTTGAATTTGAAAGTGTCTTAAATTGATTTTCAAAAGTCATATAATTTGGAACGGAGGGAGTAACAAACATCTGTCTCTTATACACATCTCCGAG	266733	G

Java script used for analyzing fastq files.

A Java script was written to provide the abundances of a certain SNP base (see Table 20-Table 23) based on the following steps:

1. Reads each sequence in a FASTQ file R1 for the sense sequence in the directory (C://projects//FASTQ//Samples).

2. Searches each sequence for “AGTGT” that is right before the N2 SNP site.

The program also contains a line for searching for all 3 SNP sites (N1, N2 and N3). These 3 positions have the sequences "GTATTT", "AGTGT" and "TATAATTT", respectively right in front of them. This line has been commented out as indicated by the double slash (//) in front of the line. This line will thus be skipped when the program is running.

3. Records the base that comes after the sequence “AGTGT” (i.e. the N2 SNP site).

4. Outputs a count of N2 SNP base in a .csv file as the base abundance.

5. If the program could not find the “AGTGT” in a specific sequence, the sequence is assigned an underscore (_) in the count.

```
package fastq;


import java.io.BufferedReader;

import java.io.File;

import java.io.FileReader;

import java.io.FileWriter;

import java.io.PrintWriter;

import java.io.IOException;

import java.time.LocalDateTime;

import java.util.Collection;

import java.util.HashMap;

import java.util.Map;


import org.apache.commons.io.FileUtils;


public class FastQCntMain {


          private static final String FASTQ_FOLDER = "C://projects//FASTQ//Samples";

          private static final String[] extentions = { "fastq" };

//        private static final String[] patterns = {"GTATTT", "AGTGT", "TATAATTT" };

          private static final String[] patterns = {"AGTGT"};

          private static final String missing = "_";


          public static void main(String[] args) {
```

```
		File directory = new File(FASTQ_FOLDER);
		Collection<File> files = FileUtils.listFiles(directory, extentions, true);
		for (File file : files)
			processFile(file);

		System.out.printf("Completed at %s", LocalDateTime.now());
	}

	private static void processFile(File file){

		System.out.printf("%s: Processing %s\n", LocalDateTime.now(), file.getName());

		String line = "";
		int lineCnt = 0;
		int cnt = 0;

		// initialize result counters
		HashMap<String, Integer> results = new HashMap<String, Integer>();

		try (BufferedReader br = new BufferedReader(new FileReader(file));
				FileWriter fileWriter = new FileWriter(file.getName() + ".exp.txt");
				PrintWriter printWriter = new PrintWriter(fileWriter))
		{
			while ((line = br.readLine()) != null) {
				cnt++;
				if (cnt == 2) {
					processLine(line, results, printWriter);
					lineCnt++;
				}
				if (cnt==4)
					cnt = 0;
			}

		} catch (IOException e) {
			e.printStackTrace();
		}

		// print results
		try (
			FileWriter fileWriter = new FileWriter(file.getName() + ".csv");
			PrintWriter printWriter = new PrintWriter(fileWriter))
			{
```

```
                    printWriter.printf("%s\n", file.getName());
                    printWriter.printf("number of sequences: %d\n", lineCnt);

                    for (Map.Entry<String, Integer> entry : results.entrySet()) {
                              String pattern = entry.getKey();
                              Integer count = entry.getValue();
                              printWriter.printf("%s,%d\n", pattern, count);
                    }

                    printWriter.close();
                    fileWriter.close();

          } catch (IOException e) {
                    e.printStackTrace();
          }
}


private static void processLine(String line, HashMap<String, Integer> results, PrintWriter printWriter) {

//                  System.out.println(line);

                    // get patterns
                    String result = "";
                    boolean notFound = false;
                    int lineLength = line.length();
                    for (String pattern : patterns) {
                              int i = line.indexOf(pattern);
                              if (i >= 0 && (i + pattern.length()) < lineLength) {
                                        int pos = i + pattern.length();
                                        result += line.substring(pos, pos+1);
                              }
                              else result += missing;
                    }

                    if (result.contains(missing))
                              printWriter.printf("%s,%s\n", line, result);

                    // store results
                    Integer cnt = results.get(result);

                    if (cnt == null)
```

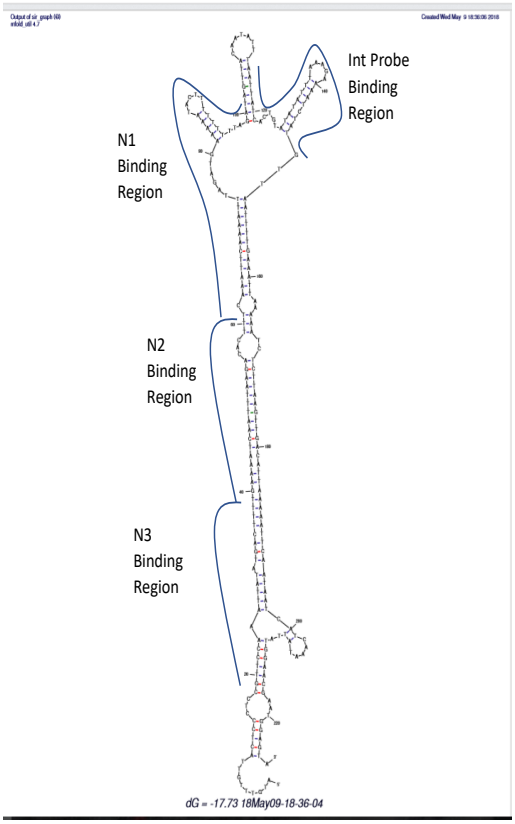
```
                    results.put(result, 1);
                else results.put(result, cnt+1);

            }
        }
```

Appendix G: Secondary Structures of 226 nt and 126 nt PCR Products

Mfold[29] was used to predict the secondary structures of the 126 nt and 226 nt antisense strands of *P. ginseng* and *P. quinquefolius* using 0.15 M Na^+ and 22°C. The structure with the most negative ΔG was selected from multiple structures that were predicted for a specific strand. Each structure has the Int. probe, N1, N2 and N3 hybridization regions marked.

(a) *P. ginseng* 226 nt Antisense PCR Strand



(b) *P. quinquefolius* 226 nt Antisense PCR Strand

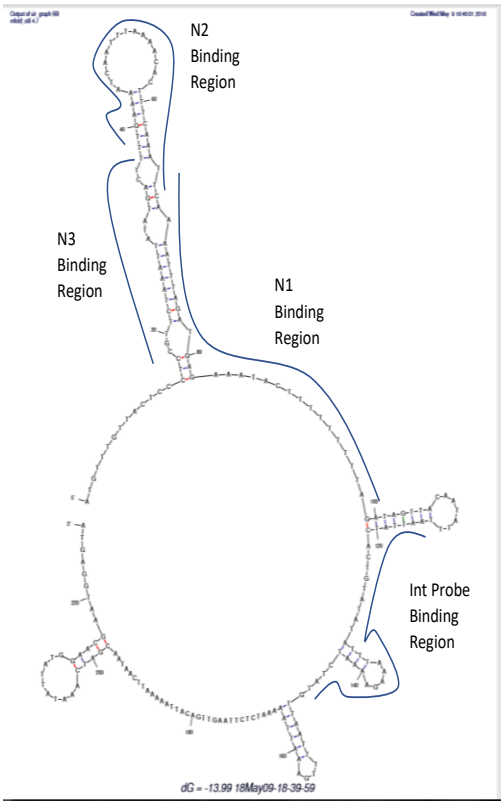


Figure 7.30: (a) The *P. ginseng* 226 nt PCR antisense strand folds into a large hairpin with few nucleotides in the N1, N2 and N3 hybridization sites exposed. These concealed sites make the hybridization reaction of the probes with the PCR strand slow at all three SNP sites, which consequently makes it better suited for SF hybridization that favors thermodynamically controlled reactions. (b) The *P. quinquefolius* 226 nt PCR antisense strand folds into a series of smaller hairpins which fully conceals only the N3 hybridization site. The partially exposed N1 and N2 SNP sites function as initiation points in the duplex formation with the respective probes. This makes these two positions more suited for dynamic hybridization which is kinetically controlled while the concealed N3 SNP sites is better suited for SF hybridization.

(a) *P. ginseng* 126 nt Antisense PCR Strand (b) *P. quinquefolius* 126 nt Antisense PCR Strand

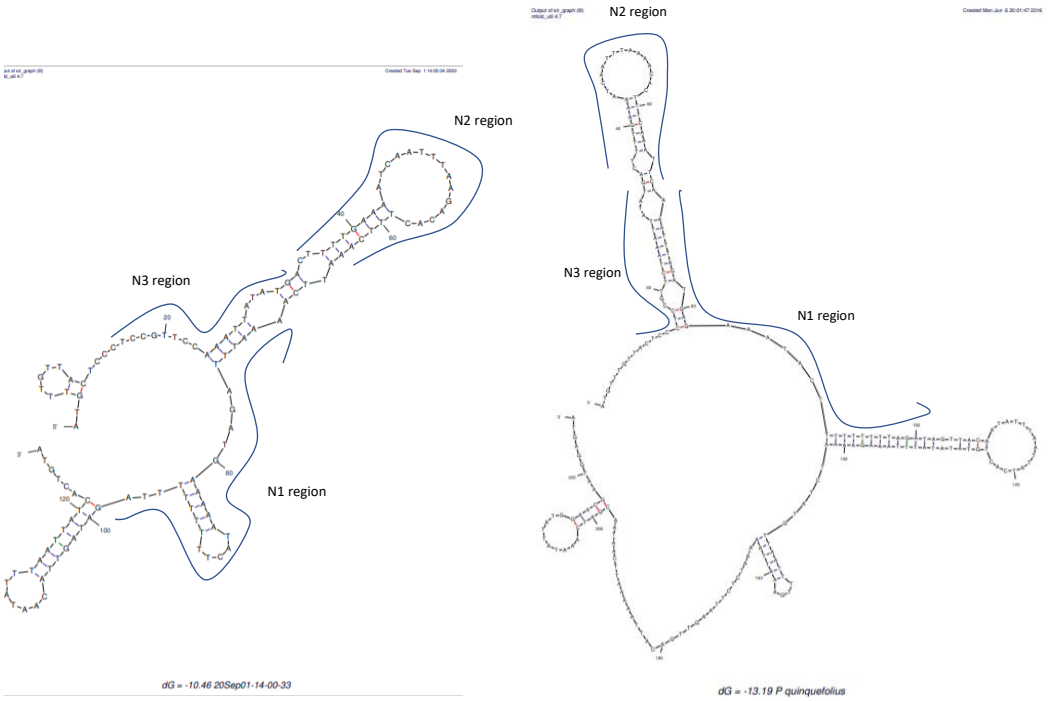


Figure 7.31: (a) The *P. ginseng* and (b) *P. quinquefolius* 126 nt PCR antisense strands fold up into similar secondary structures which have a set of exposed bases at all three SNP sites. Thus, both of these strands would be better suited for dynamic hybridization. The main difference between them is that *P. ginseng* strand has a few more bases exposed at the N3 SNP site while the *P. quinquefolius* has a few more bases exposed at the N1 SNP site.