

A practical approach to microsatellite genotyping with special reference to livestock population genetics

Compiled by

David Mburu and Olivier Hanotte

ILRI Biodiversity project

A manual prepared for the IAEA/ILRI training course on molecular characterisation of small ruminant genetic resources of Asia, October-December 2005, ILRI, Nairobi, Kenya

September 2005

Table of contents

I. Background information	4
A) What is population genetics.....	5
B) Why study genetic variation	5
C) What are the types of genetic variation.....	5
D) Molecular variation or polymorphism.....	5
E) Measures of molecular variation at individual or population levels.....	6
F) Markers for molecular variation.....	6
G) Microsatellites.....	9
H) Basic PCR.....	14
I) Primer	16
J) <i>Taq</i> DNA polymerase	20
K) “Hot start” PCR	20
L) Touch down PCR.....	21
M) PCR troubleshooting.....	22
II. Protocols	23
A) Sample collection and DNA extraction using “Magic buffer”	24
B) Reconstitution of oligonucleotides	26
C) Quantification of DNA and oligonucleotide using GeneQuant.....	28
D) Basic PCR protocol.....	29
E) Cycling profile	30
F) Goat primers -running and coloadng profile.....	31
G) Sheep primers-running and coloadng profile	32
H) Preparation of PCR Samples for 3730 and 3100 runs	33
I) ILRI segoli unit genotyping procedure	34
J) Sample sheet for 96 well plate for PCR and genotyping reaction	37
K) A 96 well sheet for genotyping run	38
L) A 48 well sheet for genotyping run.....	39
M) Genemapper	40
III. Data analysis software flow charts	46
A) Microsatellite tool kit.....	47
B) Dispan	49
C) Microsat	52
D) Phylip (neighbor.exe)	54
E) TreeView.....	56
F) Ms Dos	57
G) Displaying file extensions.....	58
H) Fstat.....	59
I) Ms Dos Genepop.....	60
J) Xlstat	62
K) WhichRun	68

L) Structure	69
M) Special characters.....	76
N) Bottleneck.....	77
O) Convert	78
P) Cervus.....	80

I. Background information

A) What is population genetics

Population genetics is about microevolution and it is the study of genetic variation in populations. This variation involves the change of allele frequencies, genotype frequencies and phenotype frequencies. Population genetics predicts diversity that is determined by a number of factors including selection, mutation rate, recombination, genetic drift and effective population size

B) Why study genetic variation

The amount and nature of genetic variation in a population allows estimates of:

- i)* Effective population size
- ii)* Population history (migration, bottleneck, recent expansion)
- iii)* Population structure
- iv)* How selection acts on genes
- v)* Location of disease genes (QTL mapping)

C) What are the types of genetic variation

- i)* Phenotypic variation (physical)
- ii)* Chromosomal variation (karyotype)
- iii)* Immunological variation (blood groups, HLA antigens)
- iv)* Protein variation (enzymes)
- v)* DNA variation (DNA sequences polymorphism)

D) Molecular variation or polymorphism

It includes:

- i)* Changes in nucleotides which could be transition or transversion. In the transition mutation, a pyrimidine (C or T) is substituted by another pyrimidine, or a purine (A or G) is substituted by another purine. The transversion mutation involves the change from a pyrimidine to a purine, or vice versa.
- ii)* Insertion or deletion of single nucleotides (indel)
- iii)* Variation in number of repeat of tandemly repeated sequences (microsatellite, minisatellite, satellite)

E) Measures of molecular variation at individual or populations levels

i) *Polymorphism*: The number of polymorphic loci or the fraction of polymorphic loci among several loci studied in a population

ii) *Heterozygosity*: The fraction of heterozygous individuals in a population maybe be estimated as:

a) Expected heterozygosity (H_E , also referred a gene diversity): This is the mean genetic diversity or the mean theoretical heterozygosity if the population is in Hardy-Weinberg equilibrium. Its calculation is based on allele frequencies. Its equal 1 minus the frequencies of homozygote

$$H_E = \sum_{i=1}^n \frac{h_i}{n} \quad \text{where } n = \text{number of loci}$$

$$h = 1 - \sum_{i=1}^n q_i^2 \quad \text{and } q_i \text{ is the } i^{\text{th}} \text{ allele}$$

b) Observed heterozygosity (H_O): Mean observed heterozygosity and can be estimated for all loci or polymorphic

$$H_O = \frac{\text{Number of heterozygous individuals}}{\text{Total number of individuals sampled}}$$

ii) Mean number of alleles (MNA): The averaged number of alleles observed in a locus

$$\text{MNA} = \frac{\text{Total number of alleles}}{\text{Number of loci}}$$

iv) *Allelic (nucleotide) diversity (A)*: One determines the number of alleles (nucleotide sequences) and their frequencies, as well as the number of nucleotide differences between each alleles pair

F) Markers for molecular variation

i) *Restriction fragment length polymorphism (RFLP)*: This technique was first developed in 1980. RFLPs are based on the analysis of patterns derived from a DNA sequence digested with known restriction enzymes. Differences are evident when the length of fragments are different, implying that the restriction enzyme cut the DNA at unrelated locations. RFLP is as a result of (a) point mutation creating or destroying a restriction site (b) insertion/deletions altering the size of a given restriction fragment. **Advantage:** (a) produces co-dominant (also known as semi-dominant) markers- this allows discrimination of homo- and heterozygotic states in diploid organisms (b) stable and reproducible - gives constant results over time, and location (c) Selective neutrality. **Disadvantage:** (a) long

methodology (b) labour intensive (c) high quality and large quantities of DNA must be used. If the amount of DNA insufficient, it may be transferred on a membrane (Southern blotting) and detection is indirect after probing with a radioactive or fluorescently labeled probe.

- ii) *PCR-RFLP* (also known as *cleaved amplified polymorphic sequence- CAPS*): Following PCR amplification of a locus, polymorphisms are detected by differences in restriction fragment sizes between individuals. **Advantage** over RFLP include (a) only little amount of DNA is required (b) short methodology (c) less labour intensive (d) probes are not required
- iii) *Randomly amplified polymorphic DNA (RAPD)*: These are short arbitrary primers (usually 10-12 base pairs in length) that bind to many places in the genome and the target product is amplified subsequently amplified by PCR. **Advantage:** (a) cost effective (b) simple and quick (c) large number of bands are produced (d) no prior sequence knowledge is necessary **Disadvantage:** (a) detection of polymorphism is limited (b) reproducibility of results maybe be inconsistent e.g. low annealing temperature may cause some unspecific non reproducible binding of primers (c) dominant markers (homozygote cannot be discerned from heterozygote so allele frequencies can not be estimated).
- iv) *Amplified fragment length polymorphism (AFLP)*: A highly sensitive method for detecting polymorphisms in DNA. Following restriction enzyme digestion of DNA, a subset of DNA fragments is selected for PCR amplification and visualisation. AFLP is widely considered as the "gold standard" for molecular epidemiological studies of pathogenic microorganisms and it is also widely used in forensic science. In plant molecular genetics, AFLP is used in phylogeny and diversity studies and in a variety of marker-assisted breeding applications. In contrast, relatively few studies have been done in animals. **Advantage:** (a) sensitive (b) large number of polymorphisms is generated (c) selective neutrality (d) highly reproducible (e) no prior sequence information or probe generation is needed. **Disadvantage:** (a) expensive (b) dominant markers (c) technically demanding
- vi) *Microsatellites*: These are DNA regions with variable numbers of short tandem repeats flanked by a unique sequence. See page 9 for more details
- vii) *Mitochondrial DNA (mtDNA)*: Mammalian mitochondrial DNA is a small (15-20 kb) circular molecule, comprising of about 37 genes coding for 22 tRNAs, two rRNAs and 13 mRNAs. Within the coding region cytochrome b is the most widely used gene for phylogenetic work. This gene evolves slowly in terms of non-synonymous substitutions. In the non-coding region the major control region for mtDNA expression is the displacement loop (D-loop), which has also been used in evolution studies. The D-loop is also

widely used in evolution studies and it has a rate of nucleotide substitution five to ten times higher than that of nuclear DNA. mtDNA polymorphisms have been widely used to investigate the structure of populations, interspecies variability, the evolutionary relationships between populations or species and for the identification of maternal lineages. Mt DNA has some importance characteristics that include maternal inheritance, high mutation rate, high copy number and lack of recombination (unlike autosomal or X chromosome specific loci). The D-loop is mainly used for intraspecies variation studies while the cytochrome b is for interspecies variation.

- ix) *Y chromosome*- The Y chromosome is the only portion of the mammalian genome that is exclusively paternally transmitted and therefore defines patrilineages. The Y chromosome consists of pseudoautosomal regions at the tips of both the long and short arm; the sex-determining gene (*SRY*) on the short arm and heterochromatin on the long arm. With exception to the pseudoautosomal region, the Y-chromosome acts as a single non-recombining unit which is male specific and effectively haploid. Y-specific single nucleotide polymorphisms have slow mutation rate (10^{-9} mutation site⁻¹ year⁻¹) and this contrasts the high rates observed in microsatellites (3×10^{-3} mutation locus⁻¹ generation⁻¹). Studies of Y chromosome polymorphism are still uncommon due to the lack of polymorphic markers in most ruminant species. In cattle Y-specific microsatellite markers have been reported and they include *INRA124*, *INRA189* and *BM861*. In horse six markers have recently been isolated (*ECAYM2*, *ECAYP9*, *ECAYH12*, *ECAYE1*, *ECAYJ10*, *ECAYA16*) Recently in sheep, a novel A/G single nucleotide polymorphic (SNP) located in the male-specific region of the Y chromosome (MSY) has been identified using comparative sequencing. A polymorphic microsatellite *SRYM18* with at least 4 alleles has also been described
- x) *Single nucleotide polymorphisms (SNPs)*: SNPs are DNA sequence variations that occur when a single nucleotide in the genome sequence is altered. They occur about once every 1 kb, within the coding and non-coding regions. For a variation to be considered a SNP, it must occur in at least 1% of the population. Most SNPs, actually about two of every three SNPs, involve the replacement of cytosine (C) with thymine (T). Many SNPs have no effect on cell function, but it is believed that they could predispose an organism to disease or influence their response to a drug. SNPs are becoming especially important as markers because (a) they are very stable, i.e. have very low mutation rates (b) they can be amplified with PCR for testing. In population genetics, it has recently been proposed that SNP analyses be used to look for footprints of adaptive evolution. Single nucleotide polymorphisms can be detected using RFLP, DNA sequencing or primer extension (“Spot-On” genotyping). The latter uses allele-specific oligonucleotides on microarrays as templates for primer extension based

genotyping. **Advantages:** (a) highly reproducible (b) very informative.
Disadvantage: (a) expensive (b) previous knowledge of sequence required

Examples of SNP genotyping methods (see details in <http://cgil.uoguelph.ca/QTL/SNPs.htm>) are:

- Single Stranded Conformation Polymorphism (SSCP)
- Allele specific oligonucleotides (ASO)
- Single nucleotide polymorphic discrimination by an electronic dot blot assay (ASO) on semiconductor microchips
- Reverse dot blot on DNA chips
- Dynamic allele specific hybridisation (DASH)
- Allele-specific PCR (=amplification refractory mutation system or ARMS test)
- Mutation detection the ARMS test in combination with the Taqman™ 5' exonuclease assay (exploiting the 5'→3' exonuclease activity of Taq DNA polymerase)
- Minisequencing and analysis of the extension products by PAGE
- Minisequencing and analysis of the extension products on DNA chips
- Minisequencing and analysis of the extension products using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDITOF)
- Pyrosequencing
- OLA
- Invasive clivage of oligonucleotide probes (Invader technology)

See table 1 for a comparison of the different genetic markers

G) Microsatellites

i) *Description:* Microratellite are short sequences of nucleotides (typically 1 to 5 bp) which are tandemly repeated. Micorsatellite alleles are characterized by different number of repeats.. Microsatellites loci are also known as: simple sequence repeats (*SSR's*), short tandem repeats (*STR's*), simple sequence tandem repeats (*SSTR*), variable number tandem repeats (*VNTR*), simple sequence length polymorphisms (*SSLP*), simple sequence repeats (*SSR*), sequence tagged microsatellites (*STMS*)

ii) *Example of microsatellites:*

a) Repeat units

AAAAAAAAAAAA = (A)₁₁ = mononucleotide (11 bp)

GTGTGTGTGTGT = (GT)₆ = dinucleotide (12 bp)

CTGCTGCTGCTG = (CTG)₄ = trinucleotide (12 bp)

ACTCACTCACTCACTC = (ACTC)₄ = tetranucleotide (16 bp)

b) Homozygous microsatellite

...CGTAGCCTTGCATCCTTCTCTCTCTCTCTCTATCGGTACTIONACGTGG... (46 bp)
...CGTAGCCTTGCATCCTTCTCTCTCTCTCTCTATCGGTACTIONACGTGG... (46 bp)

5' flanking region microsatellite locus 3' flanking region

c) Heterozygous microsatellite

...CGTAGCCTTGCATCCTTCTCTCTCTCTCTCT ATCGGTACTIONACGTGG... (46 bp)
...CGTAGCCTTGCATCCTTCTCTCTCTCTCTCTATCGGTACTIONACGTGG.. (50 bp)

5' flanking region microsatellite locus 3' flanking region

iii) Mutation

The simplest mechanism for microsatellite mutation is replication slippage with unequal crossing-over being an alternative model (<http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=hmg.section.1095>). Two main mutation models have been associated with mutation at microsatellite loci: the infinite allele model (IAM) and the stepwise mutation model (SMM). The IAM states that all mutation events will create a new allele and that new allele may have more than one repeat difference with the 'parental allele'. The SMM predicts that mutation will result in an allele which is one repeat larger or smaller. Here, difference in size conveys additional information about ancestral relationship between alleles. Under the IAM model alleles in different individual can be either "same" or "different". Under the SMM model allele can also either be the same or different but also alleles with large size differences will be more distantly related than alleles with small size difference. Wright's F-statistic underestimates population variation and is based on the IAM; Slatkin's R-statistic overestimates population variation and is based on the SMM. The average mutation rate of microsatellites ranges from 10^{-2} to 10^{-6} per locus per generation, among different species and different types of microsatellites (http://www.intl-pag.org/11/abstracts/P5o_P642_XI.html). For example a spontaneous mutation of $1.1 \pm 0.5 \times 10^{-4}$ mutations/gamete per locus has been reported in sheep (<http://www.genome.org/cgi/content/abstract/6/9/876>). In a human Brinkman *et al.* (*Am. J. Hum. Genet.*, 62:1408-1415, 1998) has reported a rate of 0 and 7×10^{-3} per locus per gamete per generation.

iv) Development of microsatellites markers

Microsatellites are developed in two ways:

- a) Use and screening of DNA repositories sequences (e.g. Genbank, EMBL)
- b) Through cloning involving generally the following step general steps (<http://www.uwyo.edu/dbmcd/molmark/lect08/lect8.html>):

- Genomic DNA extraction from tissue
- Digestion of DNA with restriction enzyme
- Insertion of the fragments into often a plasmid DNA vector
- Transfer of plasmid clones into membranes
- Probing with labeled desirable repeats tandemly repeat oligonucleoties probes
- Pick and bacterial culture of positive clones
- Extraction of plasmid DNA, and cutting of inserts with restriction enzyme
- Confirmation of the micorsatelllite nature of the insert by Southern blotting
- Sequencing of the positive clones
- Analyze of the sequence to check for "good" primer sites and useful repeat length (generally at least 8 repeats)
- Design of primer use software packages e.g.

(a) OLIGOANALYZER 3.0

<http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/Default.aspx>

(b) PRIMERQUEST

<http://www.biocompare.com/itemdetails.asp?itemid=328284>

-Screening of the loci, the quality of a genetic marker is typically measured by

- Its heterozygosity in a population
- The population information content (PIC) which equal the probability that the parent is heterozygous x probability that the offspring is informative

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^n \sum_{j=i+1}^n 2p_i^2 p_j^2$$

v) *Application*

- a) **Forensics** – individual identification and relatedness testing
- b) **Disease diagnosis** - detection of genes responsible for various genetic disorders e.g cancer
- c) **Population genetic studies**- to infer genetic diversity, population structure and relationships
- d) **Conservation biology**- identification of conservation units
- e) **Linkage analysis**-QTL mapping

vi) *Advantages*

- a) Low quantities of template DNA required (10-100 ng)
- b) High genomic abundance
- c) Random distribution throughout the genome
- d) High level of polymorphism
- e) Band profiles can be interpreted in terms of loci and alleles
- f) co-dominant markers
- g) Allele sizes can be determined with high accuracy
- h) Comparison across different gels possible using size standard
- i) High reproducibility
- j) Different microsatellites may be multiplexed in PCR or coloaded in a gel
- k) Wide range of applications
- l) Amenable to automation

vii) *Disadvantages*

- a) Initial high development costs
- b) Heterozygotes may be misclassified as homozygotes when null-alleles occur due to mutation in the primer annealing sites
- c) Stutter bands may complicate accurate scoring of polymorphisms
- d) Underlying mutation model (infinite alleles model or stepwise mutation model) largely unknown
- e) Homoplasy due to different forward and backward mutations may underestimate genetic divergence
- f) Time-consuming and expensive to develop

Table 1: Comparison of commonly used genetic markers
<http://www.fao.org/DOCREP/003/X6884E/x6884e02.htm>

Feature	RFLP	SSR	RAPD	AFLP	Isozymes	STS/EST
Origin	Anonymous/ Genic	Anonymous	Anonymous	Anonymous	Genic	Genic
Maximum theoretical number of possible loci in analysis	Limited by the restriction site (nucleotide) polymorphism (tens of thousands)	Limited by the size of genome and number of simple repeats in a genome (tens of thousands)	Limited by the size of genome, and by nucleotide polymorphism (tens of thousands)	Limited by the restriction site (nucleotide) polymorphism (tens of thousands)	Limited by the number of enzyme genes and histochemical enzyme assays available (30-50)	Limited by the number of expressed genes (10,000-30,000)
Dominance	Codominant	Codominant	Dominant	Dominant	Codominant	Codominant
Null alleles	Rare to extremely rare	Occasional to common	Not applicable (presence/absence type of detection)	Not applicable (presence/absence type of detection)	Rare	Rare
Transferability	Across genera	Within genus or species	Within species	Within species	Across families and genera	Across related species
Reproducibility	High to very high	Medium to high	Low to medium	Medium to high	Very high	High
Amount of sample required per sample	2-10 mg DNA	10-20 ng DNA	2-10 ng DNA	0.2-1 µg DNA	Several mg of tissue	10-20 ng DNA
Ease of development	Difficult	Difficult	Easy	Moderate	Moderate	Moderate
Ease of assay	Difficult	Easy to moderate	Easy to moderate	Moderate to difficult	Easy to moderate	Easy to moderate
Automation / multiplexing	Difficult	Possible	Possible	Possible	Difficult	Possible
Genome and QTL mapping potential	Good	Good	Very good	Very good	Limited	Good
Comparative mapping potential	Good	Limited	Very limited	Very limited	Excellent	Good to very good
Candidate gene mapping potential	Limited	Useless	Useless	Useless	Limited	Excellent
Potential for studying adaptive genetic variation	Limited	Limited	Limited	Limited	Good	Excellent
Development	Moderate	Expensive	Inexpensive	Moderate	Inexpensive	Expensive
Assay	Moderate	Moderate	Inexpensive	Moderate to	Inexpensive	Moderate

				expensive		
Equipment	Moderate	Moderate to expensive	Moderate	Moderate to expensive	Inexpensive	Moderate to expensive

RFLP - restriction fragment length polymorphism; SSR - simple sequence repeats (microsatellites); RAPD - random amplified polymorphic DNA; AFLP - amplified fragment length polymorphism; STS - sequence tagged site; EST - expressed sequence tags

H) Basic PCR

http://www.idtdna.com/support/technical/TechnicalBulletinPDF/A_Basic_PCR_Protocol.pdf

i) Introduction - Polymerase chain reaction (PCR) is the cardinal laboratory technology of molecular biology. Arguably one of the most powerful laboratory techniques ever discovered, PCR combines the unique attributes of being very sensitive and specific with a great degree of flexibility. With PCR it is possible to specifically address a particular DNA sequence and to amplify this sequence to extremely high copy numbers. Since its initial development in the early 1980's, dozens of variations in the basic theme of PCR have successfully been carried out. In fact, the very flexibility and application-specific variation of PCR make it seem like there are as many ways to do a PCR reaction as there are researchers doing them. Here, a basic, straight-forward PCR protocol is presented. Where appropriate, some of the choices for modifying this standard reaction that are routinely available to researchers are discussed

ii) Choosing target substrates and PCR primers - The choice of the target DNA is, of course, dictated by the specific experiment. However, one thing is common to all substrate DNAs and that is they must be as clean as possible and uncontaminated with other DNAs. Naturally, if the source material is an environmental sample such as water or soil, then the researcher must rely upon the specificity of the PCR primers to avoid amplification of the wrong product. Specificity in the choice of PCR primers should be an issue in any PCR amplification. The software OLIGOANALYZER 3.0 and PRIMERQUEST are invaluable aids both in primer design and validation. PRIMERQUEST will assist in primer design and will permit the researcher to directly assess primer specificity via a direct BLAST search of the candidate sequences. Taking candidate primer sequences into OLIGOANALYZER will allow for each primer sequence to be assessed for the presence of secondary structures whether these are hairpins or homo- and hetero-dimers

iii) Setting up the reaction - Once you have chosen the appropriate substrate and your PCR primer sequences and you have them on hand, the basic reaction components are as follows: water, 10x reaction buffer, MgCl₂, dNTPs, forward primer, reverse primer, target DNA, polymerase enzyme. One common choice available to the researcher is whether or not to use a reaction buffer that already

contains magnesium chloride. The vast majority of PCR reactions will work perfectly well at a 1.5 mM magnesium chloride concentration. For this reason PCR reaction buffers that do contain MgCl₂ are prepared so that the final concentration is 1.5 mM MgCl₂. Occasionally, however, MgCl₂ final concentrations other than 1.5 mM may be optimal. When this occurs it becomes necessary to use a reaction buffer that does not contain MgCl₂ and to add the MgCl₂ separately

iv) Choosing the reaction conditions - A PCR amplification reaction is typically characterized by a number of cycles to be run and the temperature and duration of each step in those cycles. The decision as to how many cycles to run is based upon the amount of DNA target material you start with as well as how many copies of the PCR product (amplicon) you want. In general, 25 to 35 cycles is the standard for a PCR reaction. This results in from approximately 34 million to 34 billion copies of the desired sequence using 25 cycles and 35 cycles respectively. Additional cycle numbers may be used if there is a small amount of target DNA available for the reaction. However, reactions in excess of 45 cycles are quite rare. Also, increasing the number of cycles for larger amounts of starting material is counter productive as the presence of very high concentrations of PCR product is itself inhibitory of the PCR amplification process. Once the number of cycles is selected, it is necessary to choose the temperature and duration of each step in the cycles. The first step is the DNA denaturation step where double stranded DNA become single stranded. This is routinely accomplished at 94 °C or 95 °C for 30 seconds. The second step is the primer annealing step during which the PCR primers find their complementary targets and attach themselves to those sequences. Here the choice of temperature is largely determined by the melting temperature (T_m) of the two PCR primers (OLIGOANALYZER 3.0). Again, the usual duration is 30 seconds. Finally, the last step in a PCR cycle is the polymerase extension step during which the DNA polymerase is producing a complementary copy of the target DNA strand starting from the PCR primer sequence. The usual temperature of this step is 72 °C, considered to be a good optimum temperature for thermal-stable polymerases. A common rule of thumb for the duration of this step is 30s for every 500 bases in the PCR product. However, with the increasing quality of commercially available polymerase enzymes and the associated reaction components, this time can be significantly shortened. Optimal extension time are polymerase and sequence specific. In addition to these cycling conditions, it is often desirable to add a single denaturation step of three to five minutes at 94 °C or 95 °C at the beginning of the reaction and a final extension step of a few minutes at 72 °C. A convenient shorthand way of representing a complete set of a PCR reaction conditions is: 94 °C 5:00[94 °C 0:30; 60 °C 0:45; 72 °C 2:00]35; 72 °C 7:00 which means an initial denaturing step of five minutes at 94 °C followed by 35 cycles of 94 °C for 30 seconds, 60 °C for 45 seconds and 72 °C for two minutes and then a final extension at 72 °C for seven minutes

(v) *Validating the reaction* - Once your PCR reaction has run, there are two ways of determining success or failure. The first is to simply take some of the final reaction and run it out on an agarose gel with an appropriate molecular weight marker to make sure that the reaction was successful and that the amplified product is the expected size relative to the maker. The ultimate validation of a PCR reaction is to directly sequence the amplicon product). This is often a choice that is not readily available since not everyone has access to a DNA sequencer nor will they have either the time or the funds to carry out such an analysis. One way to indirectly assess the sequence of an amplicon, however, is to carry out restriction enzyme digests on it. Given the vanishingly low likelihood that two well chosen primers will amplify an incorrect amplicon that matches the expected size, it is even more unlikely that an incorrect amplicon will give an expected pattern of restriction fragments

I) Primer

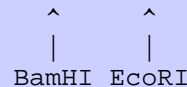
- i) *Description* - A short sequence (of RNA or DNA) from which DNA replication can initiate. May be either a synthetic DNA or RNA or a length of RNA synthesized in vivo by primase. The name primer is synonymous to oligonucleotide or oligo in short
 - ii) *Properties*
 - a) Primers must be specific of the DNA sequences to be replicated
 - b) Anneal at high temperature (typically 18 to 25 bp in length and at least 40 % GC rich)
 - c) Does not allow the formation of secondary structures that might impair the reaction (e.g. hairpin loop)
 - iii) *Primer design* –There are several softwares which could be used for primer design and details are available in the web (OLIGOANALYZER 3.0, PRIMERQUEST OLIGO 5.0). However, the basic concepts are (<http://seqcore.brcf.med.umich.edu/doc/dnaseq/primers.html>).
- (a) Sequences are always written from 5' to 3'. This includes the sequence of your template DNA (if known), the sequence of the vector DNA into which it is inserted, and the sequence of proposed primers. Don't ever write a primer sequence reversed or you will only confuse yourself and others
 - (b) Polymerase always extends the 3' end of the primer. The sequence you will read will be the same strand (sense or anti-sense) as the primer itself
 - (c) If you choose a primer sequence that you can read in your source sequence (for example, in the vector), the sequence you will obtain will extend from the primer's right (3') end

- (d) Conversely, if you choose a primer from the strand opposite to what your 'source' sequence reads, the resulting sequence will read towards the left

Examples:

Suppose you have a vector with the following sequence around the Multiple Cloning Site (the 'MCS'):

TTAGCTACTGCTTGATGCTAGTACTACATCTAGTGCTAGATGGATCCGAATTCGCTGATGCTCATATGTTAA
TAAAGAC



If you cloned your DNA of interest between the BamHI and EcoRI sites, you could sequence using the primer 'CTTGATGCTAGTACTACATC' (remember - that's written 5' to 3') and you'll obtain the following sequence from the Core:

TAGTGCTAGATG[*your-insert-'top'-strand-Bam-to-Eco*]AATTCGCTGATGC... (etc.)

What if you wanted sequence from the other strand - Eco to Bam - instead? In that case, you need to select some sequence on the *right* and then *reverse-complement* it before requesting the oligo. Picking out some sequence from the figure above:

CTGATGCTCATATGTTAATA

This is NOT the primer sequence - it is copied verbatim from the above sequence. In fact, if you used this sequence for a primer, sequencing would proceed *towards the right, away from your insert*. Instead, reverse-complement that sequence:

TATTAACATATGAGCATCAG

NOW this should produce sequence of the opposite strand:

CGAATT[*your-insert-'bottom'-strand-Eco-to-Bam*]CATCTAGCACTA... (etc.)

Some fine print: Only rarely does sequencing actually show the nucleotides immediately downstream from the primer. We have taken some didactic license in the examples above.

(iv) More advanced concepts of designing primer that works

- a) Design primers only from accurate sequence data- Sequencing has a finite probability of producing errors. Sequence obtained too far away from the primer must be considered questionable. Select a region for primer placement where the possibility of sequence error is low

b) Restrict your search to regions that best reflect your goals - You may be interested in maximizing the sequence data obtained, or you may only need to examine the sequence at a very specific location in the template. Such needs dictate very different primer placements

- Maximize sequence obtained while minimizing the potential for errors: Generally, you should design the primer as far to the 3' as you can manage so long as you have confidence in the accuracy of the sequence from which the primer is drawn. Primers on opposite strands should be placed in staggered fashion as much as possible
- Targeted sequencing of a specific region: Position the primer so the desired sequence falls in the most accurate region of the chromatogram. Sequence data is often most accurate about 80-150 nucleotides (nt) away from the primer. Do not count on seeing good sequence less than 50 nucleotides away from the primer or more than 300 (this is too short? nt away (although we often get sequence starting immediately after the primer, and we often return 700 nt of accurate sequence)

c) Locate candidate primers - Identify potential sequencing primers that produce stable base pairing with the template DNA under conditions appropriate for cycle sequencing. It is strongly suggested that you use a computer at this step. Suggested primer characteristics:

- Length should be between 18 and 30 nt, with optimal being 20-25 nt. (There are successes with primers longer than 30 and shorter than 18).
- G-C content of 40-60% is desirable.
- The T_m should be between 55°C and 75 °C. Warning: the old "4 degrees for each G-C, 2 degrees for each A-T" rule works poorly, especially for oligos shorter than 20 or longer than 25 nt. Instead, try:

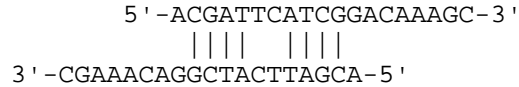
$$T_m = 81.5 + 16.6 * \log [Na] + 0.41 * (\%GC) - 675 / \text{length} - 0.65 * (\% \text{ formamide}) - (\% \text{ mismatch})$$

A simpler formula for calculation of the T_m is

$$T_m = 4(G + C) + 2(A + T)^\circ C$$

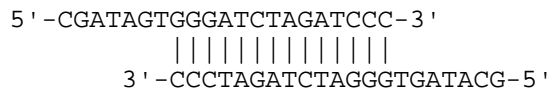
There's a web-based Tm calculator you might try at <http://www.rnature.com/oligonucleotide.html>.

- d) Discard candidate primers that show undesirable self-hybridization - Primers that can self-hybridize will be unavailable for hybridization to the template. Generally avoid primers that can form 4 or more consecutive bonds with itself, or 8 or more bonds total. Example of a marginally problematic primer:



This oligo forms a substantially stable dimer with itself, with four consecutive bonds at two places and a total of eight inter-strand bonds.

Primers with 3' ends hybridizing even transiently will become extended due to polymerase action, thus ruining the primer and generating false bands. Be somewhat more stringent in avoiding 3' dimers. For example, the following primer self-dimerizes with a perfect 3' hybridization on itself:



The above oligo is pretty bad, and almost guaranteed to cause problems. Note that the polymerase will extend the 3' end during the sequencing reaction, giving very strong sequence ACTATGC. These bands will appear at the start of your 'real' data as immense peaks, occluding the correct sequence. Most primer design programs will correctly spot such self-dimerizing primers, and will warn you to avoid them.

Note however that no computer program or rule-of-thumb assessment can accurately predict either success or failure of a primer. A primer that seems marginal may perform well, while another that appears to be flawless may not work at all. Avoid obvious problems, design the best primers you can, but in a pinch if you have few options, just try a few candidate primers, regardless of potential flaws.

- e) Verify the site-specificity of the primer - Perform a sequence homology search (e.g. dot-plot homology comparison) through all known template sequence to check for alternative priming sites. Discard any primers that display 'significant' tendency to bind to such sites. Avoid primers where alternative sites are

present with (1) more than 90% homology to the primary site or (2) more than 7 consecutive homologous nucleotides at the 3' end or (3) abundance greater than 5-fold higher than the intended priming site.

f) Choosing among candidate primers - If at this point you have several candidate primers, you might select one or a few that are more A-T rich at the 3' end. These tend to be slightly more specific in action, according to some investigators. You may want to use more than one primer, maximizing the likelihood of success. If you have no candidates that survived the criteria above, then you may be forced to relax the stringency of the selection requirements. Ultimately, the test of a good primer is only in its use, and cannot be accurately predicted by these simplistic rules-of-thumb.

J) *Taq* DNA polymerase

- i) *Description:* This is the recombinant form of the enzyme from the thermophilic eubacterium *Thermus aquaticus* BM, expressed in *E. coli*. *Taq* consists of a single polypeptide chain with a molecular weight of approximately 95 kD. It is a highly processive 5'-3' *Taq* DNA polymerase that lacks 3'-5' exonuclease activity. It is also free of nonspecific endo- and exonucleases activity and it is stable during prolonged incubations at elevated temperatures (95°C).
- ii) *Application:* a) PCR
b) RT-PCR
c) Primer extension
- iii) *Unit definition:* The concentration of the enzyme is expressed in units. One unit is defined as the amount of enzyme that incorporates 10 nmoles of dNTPs into acid-insoluble form within 30 min. at 74°C under the following conditions: 25 mM TAPS (N-tris-(hydroxymethyl)-methyl-3-amino-propanesulfonic acid, sodium salt) pH 9.3 (at 25°C); 50 mM KCl; 2 mM Mg²⁺; 1 mM β-mercaptoethanol; 200 μmol each dATP, dGTP, dTTP; 100 μM dCTP (a mix of unlabeled and [α-32P]-labeled); 12.5 μg activated salmon sperm DNA in a final volume to 50 μl.

K) “Hot start” PCR

Since components of PCR reaction are all added at room temperature or at 4 °C there is the possibility that the primers may anneal non-specifically to the template and polymerase may elongate the primer resulting in amplification that

contaminate the desired product. To prevent it a “hot start” technique is applied where essential PCR components are separated in various ways:

i) Manually: polymerase is only added to tubes after temperature is above 70°C

ii) Physical separation: using wax that melts at 75-80°C allowing mixing of all reaction components

iii) Polymerase antibodies: As temperature in tube rises the antibody is inactivated

iv) Modified polymerase: Activated at high temperatures

v) Others inhibitors: make enzyme inactive at room temperature

L) Touch down PCR

This method was first described in 1991 in order to avoid amplifying non-specific products during gene amplification. The melting point of the primer sets the upper limit on annealing temperature. At temperatures just below this point, only very specific base pairing between the primer and the template will occur. At lower temperatures, the primers bind less specifically. The earliest steps of a Touchdown PCR cycle have high annealing temperatures. For every subsequent cycle, the annealing temperature is decreased

M) PCR troubleshooting

Modified from <http://www.roche-applied-science.com/fst/amplification.htm?sis/amplification/pifs/enzymes/taq.htm>

Problem	Probable cause	Recommendation
No product	Non-optimal Mg ⁺⁺ concentration	Optimize using different Mg concentrations (0.5-5 mM)
	Amount of template in reaction is not optimal	Optimize using different template concentrations [100 - 750 ng human DNA (105 - 106 copies) per 100 µl reaction]
	Enzyme inhibitor present in reaction	Reduce or remove inhibitor (http://www.roche-applied-science.com)
	Primer annealing temperature too high or low	Optimize using different annealing temperatures
	Primer degraded or not optimal	Refer notes on the design of working primers
	Incomplete template denaturation	Reaction should reach a temperature at which complete strand separation occurs (95-97 °C)
	Machine based error	Calibrate heating block/run diagnostic program
	Mispriming	Various (http://www.roche-applied-science.com)
	NaCl concentration above 50 mM	Reduce NaCl concentration
	KCl concentration above 50 mM	Reduce KCl concentration
Misincorporation or Low Fidelity	Non-optimal Mg ⁺⁺ concentration	
	Nucleotide concentration is too high or unbalanced	Check concentration of the stock and final concentration of all nucleotides (20-200 µM of each nucleotide)
	The pH of the reaction buffer is too high	pH 8.3 is optimal
	Mispriming	
	Damaged template DNA	Check DNA integrity, minimize cycle number (25-30), reduce denature time (15-30s)
Non Specific Bands	Non-optimal Mg ⁺⁺ concentration	
	Nucleotide concentration is too high or unbalanced	
	DNA contamination / carry-over	Do PCR without adding target, used good lab practise (http://www.roche-applied-science.com)
	Primer annealing temperature is too low	Optimize anneal temp
	Mispriming	
	Primers are degraded or not optimal	
	Primer concentration is too high	Adjust the primer concentration (0.1 - 1.0 µM of each primer is optimal)
Incorrect template to enzyme ratio	Optimize template in reaction [100- 750 ng human DNA (10 ⁵ - 10 ⁶ copies) per 100 µl reaction]	
Smeared Bands	Non-optimal Mg ⁺⁺ concentration	
	Nucleotide concentration is too high or unbalanced	
	DNA contamination / carry-over	
	Primer annealing temperature is too low	
	Mispriming	
	DNase activity	Prepare fresh stock, check DNA integrity
	Oil contamination of gel sample	Spin reaction tube and remove extraction oil
Incorrect template to enzyme ratio		
Low Yield	Primer annealing temperature is too high	
	Template not clean or degraded	Purify DNA or try "Hot start" technique
	An enzyme inhibitor is present in the reaction	
	Not enough template in the reaction	
	Extension temperature too high	Use optimal temp (72 °C-1 min)
	Cycle number is too high	
	Nucleotides hydrolyzed	store nucleotide stock solutions at a concentration of at least 10 mM, 100 mM is best
	Nucleotide concentration is too high or unbalanced	
	Primer concentration is too low	Adjust the primer concentration (0.1 - 1.0 µM of each primer is optimal)
	Machine based error	
	Plateau effect	
	Evaporation	Use heated lid or 100 µl mineral oil overlay

II. Protocols

A) Sample collection and DNA extraction using “Magic buffer”

i) Sample collection

The “**Magic Buffer**” has different components: anticoagulant, anti-fungi and anti-bacteria. This buffer preserves the cell in a way that it does not break it and the DNA is kept in the nucleus where it stays stable for years. The bottle of “**Magic Buffer**” can be kept for years without losing its properties. Keep it in a dry and temperature stable room

- Collect the sample by veno-puncture with an **EMPTY** 10 ml Venoject or Vacutainer tube (*red rubber*) or with an **EDTA** 10 ml Venoject or Vacutainer tube (*purple rubber*)
- Stop taking the sample when volume equals 5 ml
- IMMEDIATELY** add 5 ml of “**Magic Buffer**” with a syringe.
- Mix by inverting the tube gently
- Write the ID number of the individual

The blood samples mixed with “**Magic Buffer**” should be kept at 4-6°C (fridge temperature). **NEVER FREEZE** the sample. You can also keep it at room temperature avoiding temperatures above 30°C or direct sun on them but it is recommended this only if you will use it quite soon and are not interested in keeping it. You will see that with time that blood, which is highly red, turns very dark. This is not a problem, the samples only suffer an oxidation but this does not affect the cells or the DNA. You will see that with time, the cells will be at the bottom of the tube and the rest of the volume will be a reddish transparent liquid. This is not a problem, if you have mixed gently by inverting the tube several times when sampling

ii) When you have to extract DNA you have to do the following

- Take out the tube you want to DNA extract from the fridge the day before and leave it at room temperature
- Next day, mix gently the tube by inverting it to homogenize the content
- Take a sample of the mixed solution by opening the rubber end
- Store the tube again at 4-6°C

iii) DNA extraction – Salting-out protocol

- Take 700 µl the sample of blood + “**Magic Buffer**”
- Add the same volume of **autoclaved distilled water** vortex and centrifuge at 10.000g 10'. The objective of this step is the lyses of red cells
- Discard the supernatant and add the same volume of **autoclaved distilled water**
- Vortex until the pellet is disaggregated and centrifuge at 10.000g 10'
- Discard the supernatant
- Add 200 µl **lyses buffer (160mM sucrose, 80mM EDTA (pH 8.0), 100mM Tris (pH 8.0), 0,5% SDS)** and 2 µl **proteinase K solution** to the white cells pellet and homogenize

by inverting the tubes several times. If the pellet remains fixed to the bottom of the tube crunch it with the pipette-tip

-Incubate at 58°C for 4 hours, observing the disappearance of any pellet and debris.

-Add 100 µl of **NaCl 4.5M**

-Mix thoroughly by inversion until homogenisation

-Add 225µl of **chloroform**

-Mix by shaking for 10 min

-Centrifuge for 10 min at 14.000g

-Transfer upper phase in new 1,5 ml tube (~200µl) making sure not to touch the interphase

-Add an equal volume of **isopropanol** (~200µl) in order to precipitate the DNA of the aqueous solution. Observe the formation of white clouds in the solution

-Shake thoroughly by inversion or vortex

-Immediately centrifuge for 10 min at 14.000 g, orienting the tubes in the microcentrifuge, in order to know exactly the position of the pellet at the bottom of the tube

- After centrifugation, check for the presence of the pellet at the bottom. Sometimes the pellet can stick to the tube lateral wall or cover. In this case, shake and centrifuge it again

-Remove all the supernatant carefully, observing the position and characteristic of the DNA with the pellet and avoiding touching it

-Wash the pellet with 500 µl **ethanol 70%**, by adding the alcoholic solution carefully.

-Incubate for 15 min at room temperature

-Centrifuge for 10 min at 14.000 g

-Decant the alcoholic solution or remove it using a micropipette. Keep the visual control that the pellet stays where it should be

-Dry at room temperature or for 5 min at a vacuum centrifuge

-Add 100 µl of **autoclaved distilled water** to the dried DNA pellet

-Incubate in water bath (37°C overnight or 56°C for 4 to 6 hours)

-Using a micropipette, homogenise the solution by pipetting in and out about 30 times

-Quantify the DNA either by 2% agarose gel electrophoresis or by spectrophotometer

-Store at -20°C

B) Reconstitution of oligonucleotides

i) Moles and Molar

NB: Note the meaning and symbols of two important terminologies (a) Mole and (b) Molar (also known as molarity) to avoid confusion

Mole = absolute amount of a substance (Symbol = mol)

Molar or Molarity = Concentration of a substance in a liquid (Symbol = mol/l or M)

(a) 1 mol/l = a molar solution = a solution with a molarity of one

(b) A solution containing a total of 4 pmol (**picomoles**) of primer in a PCR reaction of 10 μ l will have a final primer concentration of 400 nM (**nano Molar**). Calculations are follows:

4 pmol in 10 μ l = 400 pmol in 1 ml = 400,000 pmol per litre = 400 nmol/l
Since mol/l is the same as M: Then 400 nmol/l = 400 nM.

ii) New capillary based primers from applied biosystems

These primers are synthesized on a scale of 0.2 μ M DNA and are supplied as a lyophilized powder of 80 nmol (confirm on despatch note).

a) Dilute the primer in 100 μ l of TE buffer (1 mM Tris-HCL pH8.0/0.01 mM EDTA) to give a concentration of 80 nmol in 100 μ l = 800 pmol/ μ l. This should be stored as the stock at -20 $^{\circ}$ C.

b) Prepare a working dilution of 2 pmol/ μ l by diluting 1 part of the stock (800 pmol/ μ l) with 399 μ l of sterile water and store at -20 $^{\circ}$ C.

NB: A good number of 10 μ l PCR reactions (protocols vary) use between 4 to 5 pmol of each primer which is equivalent to 400 nM and 500 nM, respectively.

Make sure that all fluorescently labeled oligos are stored away from natural daylight

iii) Primers internally synthesized by ILRI SEGOLI unit

- a) The primers are supplied in a lyophilized form
- b) Dilute the primer (1:100) in TE buffer and quantify with GeneQuant *pro* using pmol/ μ l as the units.
- c) Dilute as necessary to a working concentration of 2 pmol/ μ l using sterile water
- d) If you already have primer that is quantified in μ g you can convert it to pmole using the formulae below:

$$\frac{\mu\text{g} \times 1\,000\,000}{\text{primer length} \times 325^*}$$

e.g. 0.1 μ g of a 20 mer

$$(0.1 \mu\text{g} \times 1\,000\,000)/(20 \times 325) = 15.4 \text{ pmoles primer}$$

- e) Likewise equation for conversion of pmoles of primer to μ g of primer is as follows:

$$\frac{\text{pmole} \times \text{primer length} \times 325^*}{1\,000\,000}$$

e.g. 10 pmoles of a 25 mer

$$(10 \times 25 \times 325)/ 1\,000\,000 = 0.081 \mu\text{g primer}$$

* Average molecular weight of a nucleotide

C) Quantification of DNA and oligonucleotide using GeneQuant

i) Setting up the spectrophotometer for DNA quantification

Press “**Set-up**” key followed by “**enter**” key

Type in the **Month** and press “**enter**” key

Type in the **Date** and press “**enter**” key

Type in sample **number** and press “**enter**” key

Select printer **on** and press “**enter**” key

Select Autoprint **on** and press “**enter**” key

Select serial **off** and press “**enter**” key

Press “**STOP**” key

Press “**Set-up**” key followed by “**enter**” key

Press “**DNA or RNA**” key and press “**enter**” key*

Enter path length **10** and press “**enter**” key

Select units e.g. **µg/ml** and press “**enter**” key

Select “**Yes**” for 230 nm reading and press “**enter**” key*²

Select “**Yes**” for 320 nm reading and press “**enter**” key

Enter the **dilution factor** and press “**enter**” key*³

Press “**STOP**” key

ii) Setting up the spectrophotometer for oligonucleotide quantification

Repeat same procedure as in (A) but Press **Oligo** in step (*). Also select pmol/ul for units (*²) and select use default factor 33 after step (*²) before pressing “STOP” key

NB: The geneQuant has now been connected to a computer and a QuantMant software for data capture has been developed-Refer QuantMant manual for details

D) Basic PCR protocol

- 1) Wipe down the bench area
- 2) Take the PCR components out of the freezer to thaw – place them into a tube rack to thaw, and then into a "chiller box" to stay cold while assembling your PCR. The components are: water, 10X buffer, dNTPs, MgCl₂, and primers
- 3) Take DNA extracts out of freezer (or fridge)
- 4) While components are thawing:

Fill out a PCR worksheet with the sample identifiers of all samples to be used. Following the master mix template sheet, calculate the amount of each component needed for the total number of samples + 2. Remember to include a tube at the end for the "negative control" (a tube with no template added to check for contamination)
- 5) Transfer template 1 µl DNA (40 ng) into the respective wells of the microtitre plate
- 6) When all the components have thawed, prepare the Master Mix (MM) in a 1.5 ml tube, adding each component in the order listed on the PCR form: water, 10x buffer, dNTPs, MgCl₂, then primers. **Be sure to mix all tubes well** (by shaking , then shake contents down into the bottom of the tube) before pipetting them into the MM
- 7) Take out *Taq* from the freezer into the MM then close the MM tube. It is not necessary to shake up the *Taq* before pipetting, but you may want to spin down the *taq* tube to force contents to the bottom. Replace the *taq* into freezer immediately after use
- 8) Mix the MM solution by pipetting
- 9) Distribute the MM (9 µl) into your labelled microtitre plate
- 10) Place the plate into the thermal cycler and start the program
- 11) After the PCR is complete, run out samples on a mini-gel (separate protocol) or store in the PCR refrigerator until the minigel is run

E) Cycling profile

i) Basic PCR

Activity	Temperature	Duration	# Cycles
Denature	95 °C	5 min	1
Denature	95 °C	30s	} 35
Anneal	48 - 65 °C	1 min	
Extension	72°C	1 min	
Extension	72°C	10 min	1
Hold	15 °C	-	-

ii) Touch down PCR

Activity	Temperature	Duration	# Cycles
Denature	95 °C	5 min	1
Denature	95 °C	30 s	} 3
Anneal	55 °C	30 s	
Extension	72 °C	30 s	
Denature	95 °C	30 s	} 3
Anneal	52 °C	30 s	
Extension	72 °C	30 s	
Denature	95 °C	30 s	} 20
Anneal	49 °C	30 s	
Extension	72 °C	30 s	
Denature	95 °C	30 s	} 10
Anneal	46 °C	30 s	
Extension	72 °C	30 s	
Extension	72 °C	10 min	1
Hold	15 °C	Hold	-

F) Goat primers -running and coloding profile

	Marker	Anneal	Mg ⁺⁺	PCR	Range (bp)	Dye	Dilution*	Coloding Block
1	ILSTS029	55°C	2.0mM	Basic	135-185	Ned	1 in 16	1
2	BMS1494	55°C	2.0mM	Basic	235-300	6Fam	1 in 32	
3	MAF035	55°C	2.0mM	Basic	90-130	Pet	1 in 32	
4	SRCRSP3	55°C	2.0mM	Basic	95-135	Ned	1 in 32	2
5	BM1818	55°C	2.0mM	Basic	240-310	Vic	1 in 32	
6	OARFCB20	55°C	2.0mM	Basic	80-130	6Fam	1 in 64	
7	SRCRSP7	55°C	2.0 mM	Basic	110-140	Pet	1 in 4	3
8	OARAE54	55°C	2.0mM	Basic	105-145	Ned	1 in 1	
9	ILSTS005	55°C	2.0mM	Basic	160-230	6Fam	1 in 64	
12	SPS113	58°C	2.0mM	Basic	125-170	Pet	1 in 16	4
13	CSR247	58°C	2.0mM	Basic	210-260	Vic	1 in 16	
14	INRA0132	58°C	1.5mM	Basic	125-175	6Fam	1 in 16	
15	INRA023	58°C	2.0mM	Basic	180-240	Pet	1 in 4	5
16	MAF209	58°C	2.0 mM	Basic	95-150	Vic	1 in 128	
17	OARFCB48	58°C	2.0mM	Basic	140-180	6Fam	1 in 4	
18	MCM527	58°C	2.0mM	Basic	155-195	Ned	1 in 4	6
19	SRCRSP5	58°C	2.0mM	Basic	130-200	Vic	1 in 128	
20	ILSTS087	58°C	2.0mM	Basic	120-190	6Fam	1 in 16	
10	SRCRSP9	55°C	2.0mM	Basic	80-150	6Fam	1 in 16	7
11	TGLA53	55°C	2.0mM	Basic	110-170	Vic	1 in 32	
25	MAF70	65°C	2.0mM	Basic	120-190	Pet	1 in 16	
21	ILSTS11 (D14S16)	58°C	2.0mM	Basic	250-300	6Fam	1 in 16	8
22	SRYM18	58°C	1.0mM	Basic	80-170	6Fam	1 in 16	
23	INRA063	53°C	2.0mM	Basic	145-195	Vic	1 in 64	
24	ETH10	53°C	1.5 mM	Basic	190-220	Vic	1 in 16	9
26	BM6444	65°C	2.0mM	Basic	110-210	Pet	1 in 16	
28	MAF065	Touch down	1.5mM	Touch down	100-160	Ned	1 in 8	

* individual dilutions

When coloding like three dyes use 1:3 dilution (1 part PCR + 2 parts DiFI/Liz)

G) Sheep primers-running and coloding profile

	Marker	Annealing	Mg ⁺⁺	PCR	Range (bp)	Dye	Dilution*	Coloding Block
1	OARVH72	55°C	2.0mM	Basic	110-150	Ned	1 in 128	1
2	SRCRSP5	55°C	2.0mM	Basic	130-160	Vic	1 in 128	
3	OARFCB20	55°C	2.0mM	Basic	80-130	6Fam	1 in 128	
4	BM 8125	55°C	1.5mM	Basic	110-130	Ned	1 in 256	2
5	MCM 42	55°C	2.0mM	Basic	80-120	Pet	1 in 16	
6	TGLA53	55°C	2.0mM	Basic	110-175	Vic	1 in 4	
7	OAFCEB128	55°C	2.5 Mm	Basic	80-150	6Fam	1 in 1	
8	OARFCB304	55°C	2.0 mM	Basic	140-200	Vic	1 in 8	3
9	OARHH47	55°C	4.5mM	Basic	110-170	Ned	1 in 4	
10	SRCRSP9	55°C	2.0 mM	Basic	80-150	6Fam	1 in 32	
11	DYMS1	58°C	2.0 mM	Basic	140-230	Ned	1 in 32	4
12	OARCB226	58°C	2.0 mM	Basic	100-170	Vic	1 in 64	
13	ILSTS011	58°C	2.0mM	Basic	250-300	6Fam	1 in 8	
14	MAF209	58°C	2.0 mM	Basic	95-150	Vic	1 in 128	5
15	BM 1329	58°C	1.5mM	Basic	140-190	6Fam	1 in 4	
16	ILSTS49	58°C	4.5 mM	Basic	120-200	Vic	1 in 4	
17	OARJMP29	63°C	1.5mM	Basic	80-170	Pet	1 in 64	6
18	OARFCB 11	63°C	4.5mM	Basic	100-160	Vic	1 in 16	
19	OARFCB129	63°C	1.5mM	Basic	100-150	Ned	1 in 4	
20	OARCP34	48°C	4.5mM	Basic	100-160	Pet	1 in 2	7
21	HUJ616	48°C	2.5 Mm	Basic	100-170	Pet	1 in 2	
22	HSC	48°C	2.0mM	Basic	250-320	6Fam	1 in 8	
23	MCM 527	50°C	1.5mM	Basic	155-195	Ned	1 in 32	8
24	ILSTS44	50°C	1.5mM	Basic	155-175	Pet	1 in 8	
25	OARCP38	53°C	2.5 Mm	Basic	117-140	Vic	1 in 2	
26	MAF214	65°C	1.0mM	Basic	160-200	6Fam	1 in 128	9
27	OAER129	Touch down	4.5 mM	Touch down	12-180	6Fam	1in 8	
28	MAF065	Touch down	1.5mM	Touch down	100-160	Ned	1 in 8	
29	SRYM18	60°C	1.5 mM	Basic	80-170	6Fam	1 in 32	Y chromosome-male only

* individual dilutions

When coloding like three dyes use 1:3 dilution (1 part PCR + 2 parts DiFI/Liz)

H) Preparation of PCR Samples for 3730 and 3100 runs

- 1) Amplify your target product using primer pairs of which the forward is labelled with a capillary based dye: 6FAM (Blue), PET (Red), VIC (Green) and NED (Yellow)
- 2) Dilute PCR product in triple distilled water (e.g. 1:16 dilution – varies with individual markers)*
- 3) Prepare internal standard by adding 10 µl of LIZ (Orange) standard (stored at 4 °C) to 1 ml of HiDi[®] formamide (stored in aliquots of 1 ml at -20 °C) and mix by pipeting
- 4) Pipette 1 µl of dilute PCR product into individual wells of the microtitre plate
- 5) Transfer 9 µl of the standard/formamide mix into individual wells (already has 1 µl of diluted PCR product) of a microtitre plate and mix by pipeting
- 6) Prepare sample sheet and mail it to SEGOLI unit
- 7) Hold the processed samples in the fridge at SEGOLI unit

*** Primers should be optimized by running different dilutions of the PCR product and choosing the one giving the best signal profile (signal/noise ratio and RFU). An example of a dilution schedule is: 1:1, 1:2, 1:4, 1:8, 1:16, and 1:32**

NB: Always wear powder free hand gloves. Powder/proteins in samples will reduce the lifespan of the capillaries/array.

I) ILRI segoli unit genotyping procedure

i) Fragment Analysis Samples

Samples in a 96-well or 384-well plate should be left in the Fridge in the Segoli Facility Lab 4 Bay A. The completed sample sheet should be sent via e-mail to SegoliLab@cgiar.org. Upon completion of the run the data should ideally be posted to a folder on a server where access will be granted to the originators. Due to the shortage of computers that can run the GeneMapper software the data will be posted to one of the computers in the Segoli Lab and the originators informed.

ii) Submitting Samples for Fragment Analysis on the 3730

If submitting samples in a 96 well PCR plate please follow the following instructions. Check with Segoli Unit staff about compatibility of your PCR plate with the capillary instrument. The GA 3730 Instrument loads and runs the first 48 samples from odd numbered columns (purple columns), then loads and runs samples from even numbered columns (white background). See the illustration below:

A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12
B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12
C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12
D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12
E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	E12
F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12
G01	G02	G03	G04	G05	G06	G07	G08	G09	G10	G11	G12
H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12

If you have less than 48 samples we suggest that you arrange your samples in the purple colored wells. Kindly use the [3730_48Template](#) file to fill in your sample sheet and input sample name in the corresponding well position. Please fill in all shaded text. The PlateName should match the label on the plate left in the Segoli Unit Fridge. The instrument will load and run the plate once for samples arranged this way. The Segoli Unit will charge US\$38 for these samples.

For 49 - 96 samples, we request that you input your sample name in the corresponding well position in the [3730_96Template](#) file. Please fill in all shaded. The PlateName should match the label on the plate left in the Segoli Unit Fridge. The instrument has to load and run twice to process all the samples. Please contact the Segoli Unit staff if you have any problems creating the sample sheets.

iii) Submitting Samples for Fragment Analysis on the 3100

If submitting samples in a 96 well PCR plate please follow the following instructions. Check with Segoli Unit staff about compatibility of your PCR plate with the capillary instrument. The GA 3100 Instrument loads and runs the first 16 samples from two adjacent columns (purple columns), then loads and runs samples the next two adjacent columns (white background) and so on. See the illustration below:

A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12
B01	B02	B03	B04	B05	B06	B07	B08	B09	B10	B11	B12
C01	C02	C03	C04	C05	C06	C07	C08	C09	C10	C11	C12
D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12
E01	E02	E03	E04	E05	E06	E07	E08	E09	E10	E11	E12
F01	F02	F03	F04	F05	F06	F07	F08	F09	F10	F11	F12
G01	G02	G03	G04	G05	G06	G07	G08	G09	G10	G11	G12
H01	H02	H03	H04	H05	H06	H07	H08	H09	H10	H11	H12

If you have less than 16 samples we suggest that you arrange your samples in the purple colored wells. Kindly use the file [3100FA_SampleSheet](#) to fill in your samples sheet and input sample name in the corresponding well position. Please fill in all shaded. The PlateName should match the label on the plate left in the Segoli Unit Fridge. The instrument will load and run the plate once for samples arranged this way.

If you have more than 16 samples use the correct sheet tab from [3100FA_SampleSheet](#) that will correspond with the number of samples you need to run. We request that you input your sample name in the corresponding well position in the file. Please fill in all shaded. The PlateName should match the label on the plate left in the Segoli Unit Fridge. The instrument has to load and run multiples of 16 times depending on the number of your samples. The price will depend accordingly. Please contact the Segoli Unit staff if you have any problems creating the sample sheets.

iv) Filter Set G5 for Fragment Analysis

The Capillary Instruments use the G5 filter set:

Color	Dye	Absorption Max (nm)	Emission Max (nm)
Blue	6-FAM	494	522
Green	VIC	538	554
Yellow	NED	546	575
Red	PET	558	595

Color	Dye	Absorption Max (nm)	Emission Max (nm)
Orange	LIZ	638	655

Four dyes, therefore, can be used to label PCR fragments and LIZ will be used as the internal size standard

v) Consumables available from the SegoliLab

The following consumables are available from the Segoli Lab on a cost recovery basis. We do, however, send out an order to Applied BioSystems quarterly. If you foresee a requirement from this vendor please let us know in advance so that we may consolidate it with other the requests

ABI PRISM 384-Well Clear Optical Reaction Plate with Barcode P/N 4309849

ABI PRISM 96-Well Reaction Plate P/N N8010560

GeneScan < -500 LIZ Size Standard P/N 4322682

Hi-Di Formamide Genetic Analysis Grade P/N 4311320

J) Sample sheet for 96 well plate for PCR and genotyping reaction

Experiment: _____

Primer: _____

































































































Plate ID: _____

Anneal temp: _____

Date: _____

PCR type: _____

Mg⁺⁺: _____

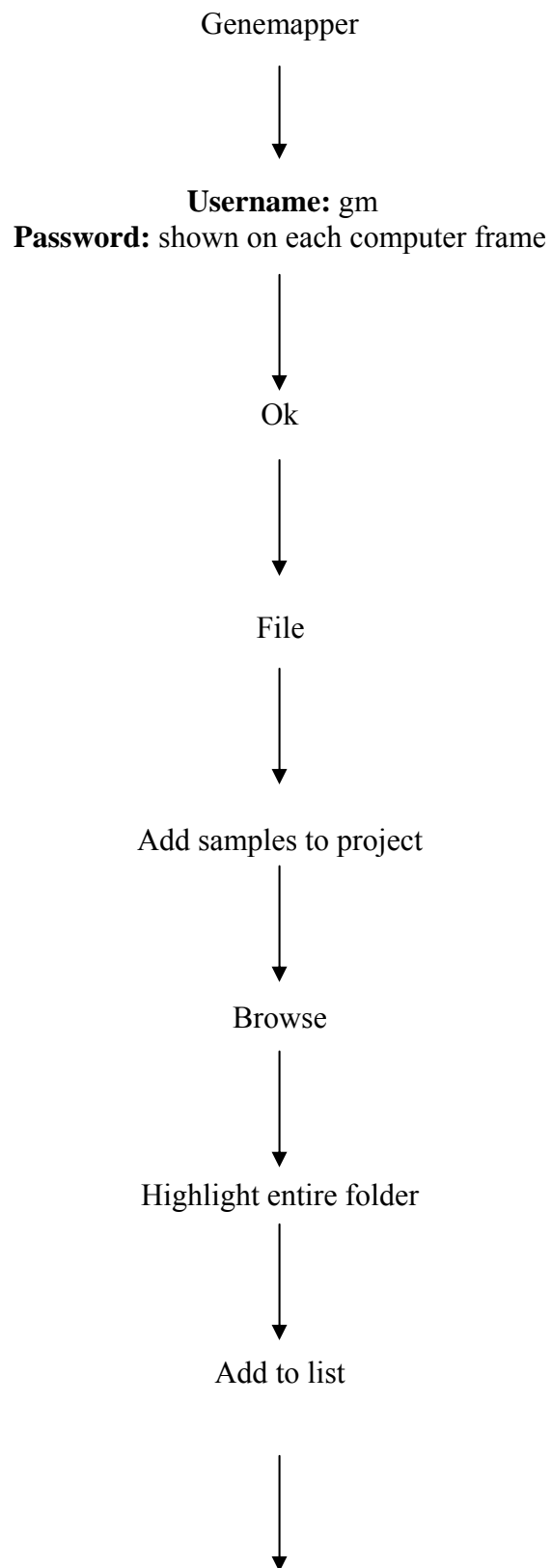
	1	2	3	4	5	6	7	8	9	10	11	12
A												
B												
C												
D												
E												
F												
G												
H												

L) A 48 well sheet for genotyping run

NB: For a 48-well either use the odd rows or the even rows and delete the others. Please fill in everything that is shaded blue and remember to re-label the sheet with your plate name. Then save and send SegoliLab@cgiar.org a copy of the text delimited file.

Container Name	Plate ID	Description	Container/T	App/T	Owner	Operator	Plate Seal	Scheduling	Pref							
ASO2	ASO2		96-Well	Regular	Mburu	Mburu	Septa	1234								
AppServer	Appliance															
GeneMapper	GeneMapper_ea3e5c24baf911d9bc2500d56296c2e															
Well	Sample Name	Comment	Sample Ty	Sn	Set	Analysis Method	Panel	User-Defin	Size	Stand	User-Defin	User-Defin	Results	Group 1	Instrument	Protocol
A01	MAF214s1		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B01	MAF214s2		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C01	MAF214s3		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D01	MAF214s4		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E01	MAF214s5		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F01	MAF214s6		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G01	MAF214s7		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H01	MAF214s8		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H02			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A03	MAF214s9		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B03	MAF214s10		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C03	MAF214s11		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D03	MAF214s12		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E03	MAF214s13		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F03	MAF214s14		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G03	MAF214s15		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H03	MAF214s16		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H04			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A05	MAF214s17		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B05	MAF214s18		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C05	MAF214s19		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D05	MAF214s20		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E05	MAF214s21		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F05	MAF214s22		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G05	MAF214s23		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H05	MAF214s24		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H06			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A07	MAF214s25		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B07	MAF214s26		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C07	MAF214s27		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D07	MAF214s28		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E07	MAF214s29		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F07	MAF214s30		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G07	MAF214s31		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H07	MAF214s32		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H08			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A09	MAF214s33		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B09	MAF214s34		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C09	MAF214s35		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D09	MAF214s36		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E09	MAF214s37		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F09	MAF214s38		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G09	MAF214s39		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H09	MAF214s40		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H10			Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
A11	MAF214s41		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
B11	MAF214s42		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
C11	MAF214s43		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
D11	MAF214s44		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
E11	MAF214s45		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
F11	MAF214s46		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
G11	MAF214s47		Sample			Microsatellite Default	None	GS500LIZ_3730		Asian_Sheep					GMG5DS33default	
H11	MAF214s48		Sample			Microsatellite Default	None	GS500LIZ_37								

M) Genemapper



Add



Okay



Analyse

(make sure it is microsatellite default)



Project name (e.g. ASO2-23-07-2005)



Edit



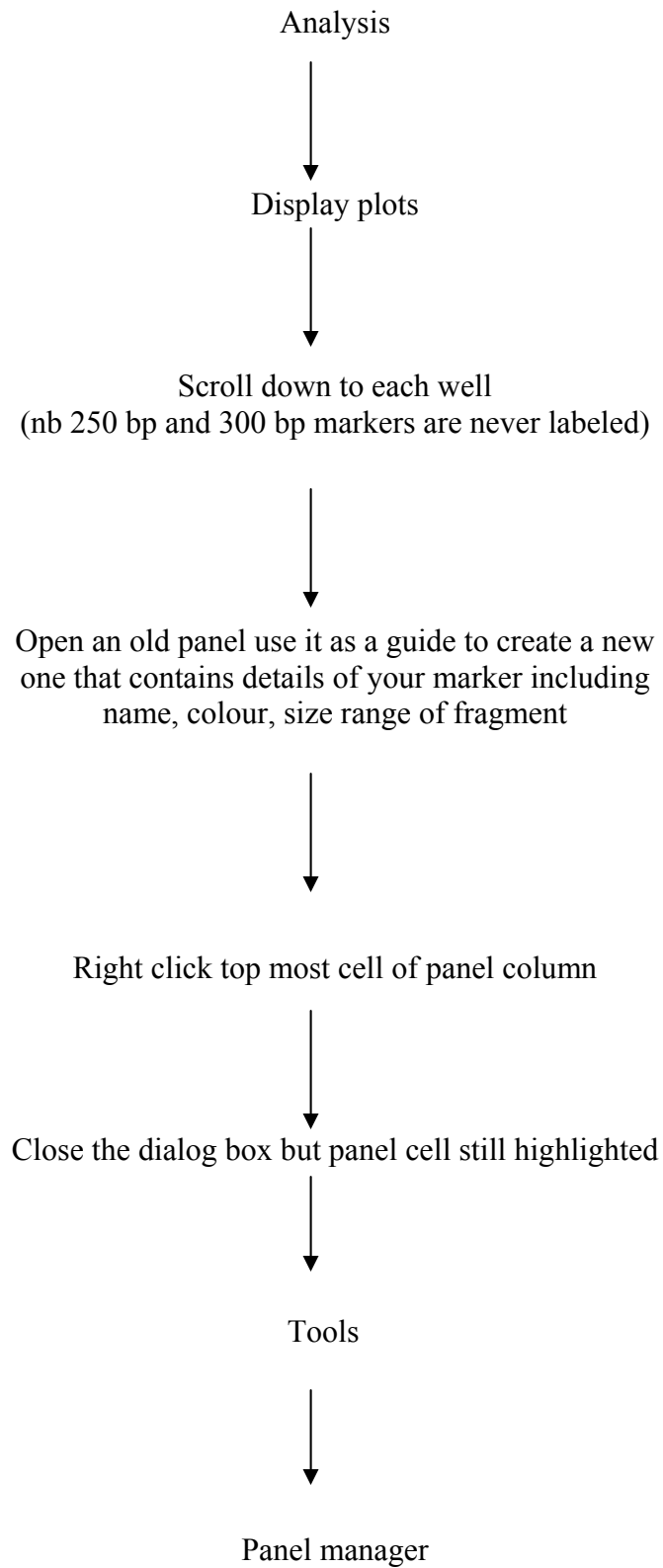
Sort

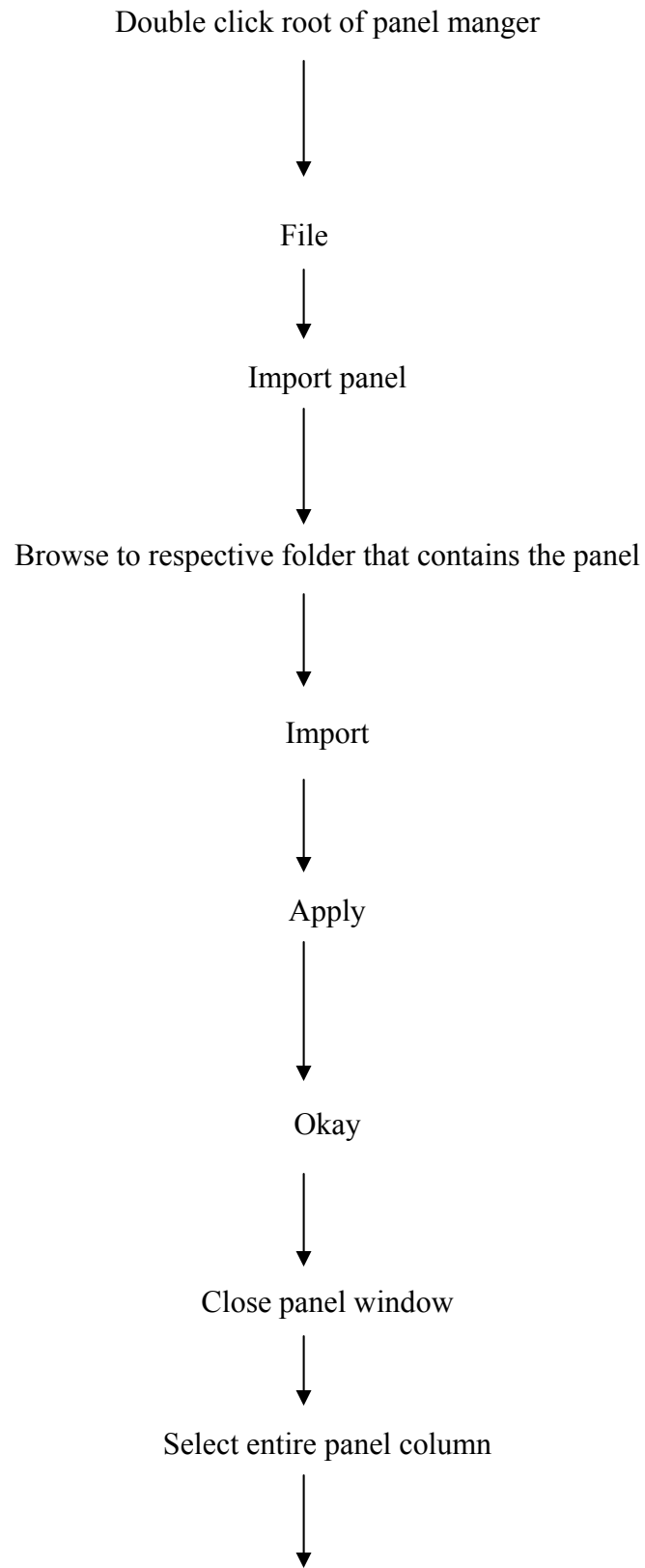


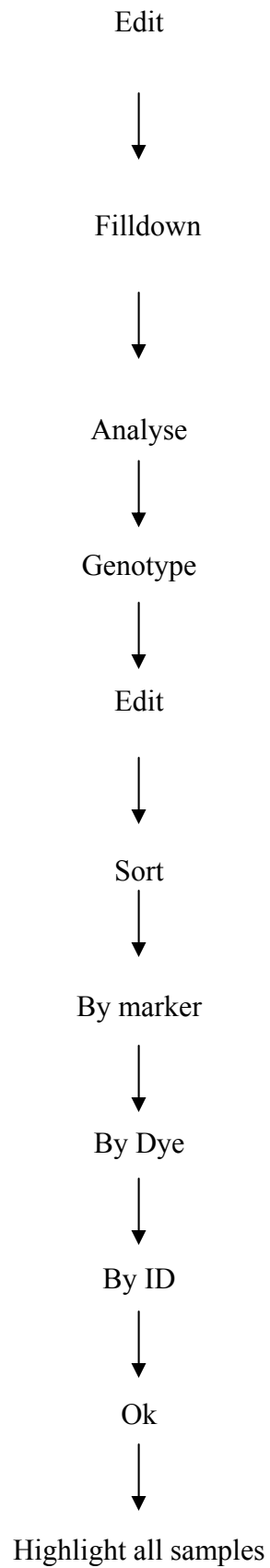
By sample name



Highlight entire column (e.g. for sample name)







Display plot
(good RFU are between 2500 to 5000 units)



File



Export table



Browse (identify folder to host file)



Export



Table is made in tab delimited



Open table by excel

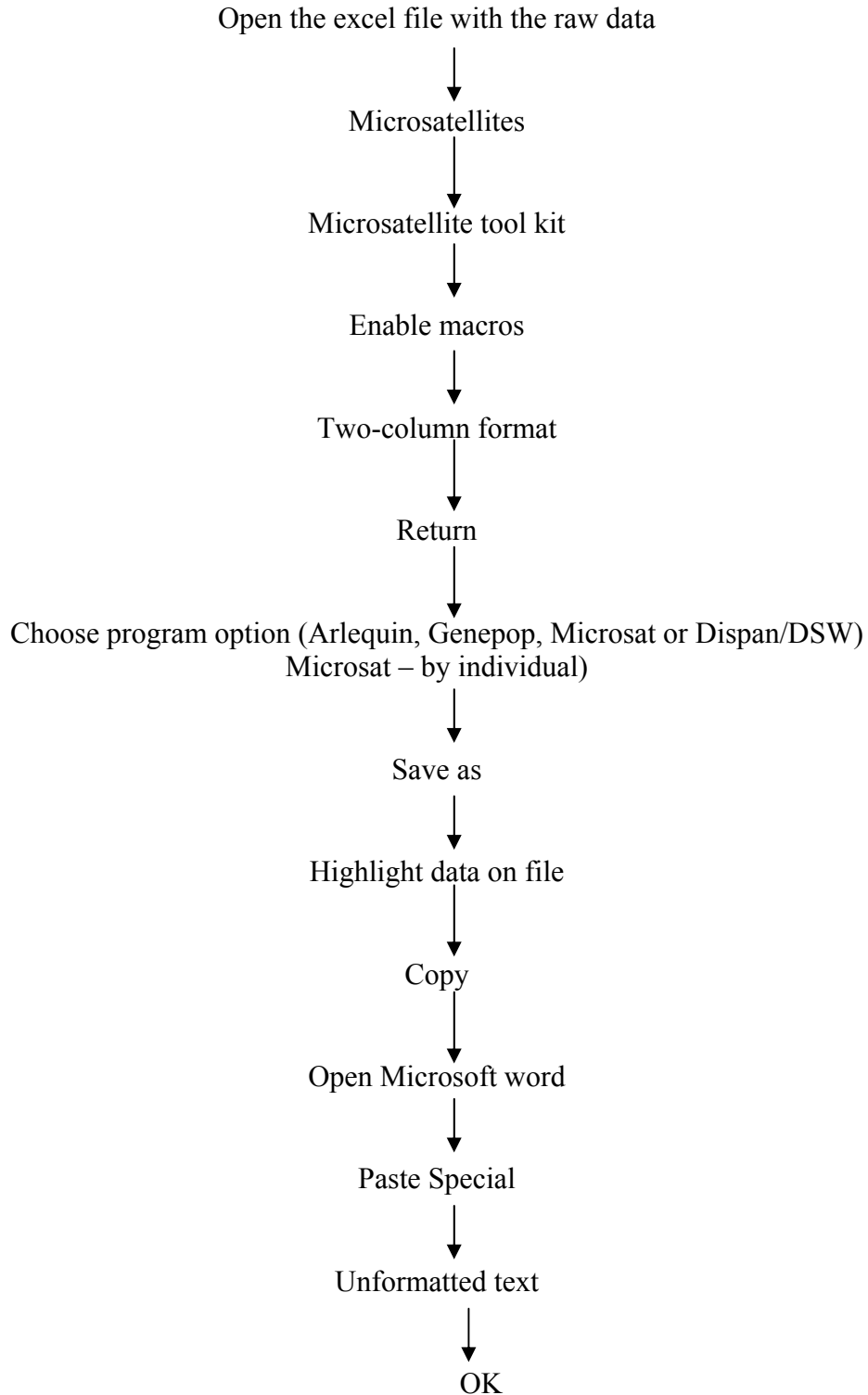


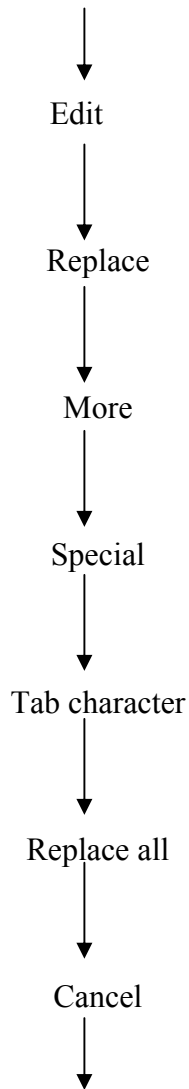
Edit table to remove the unwanted columns

III. Data analysis software flow charts

A) Microsatellite tool kit

Before data is analysed by various softwares it must be converted into the appropriate format as follows

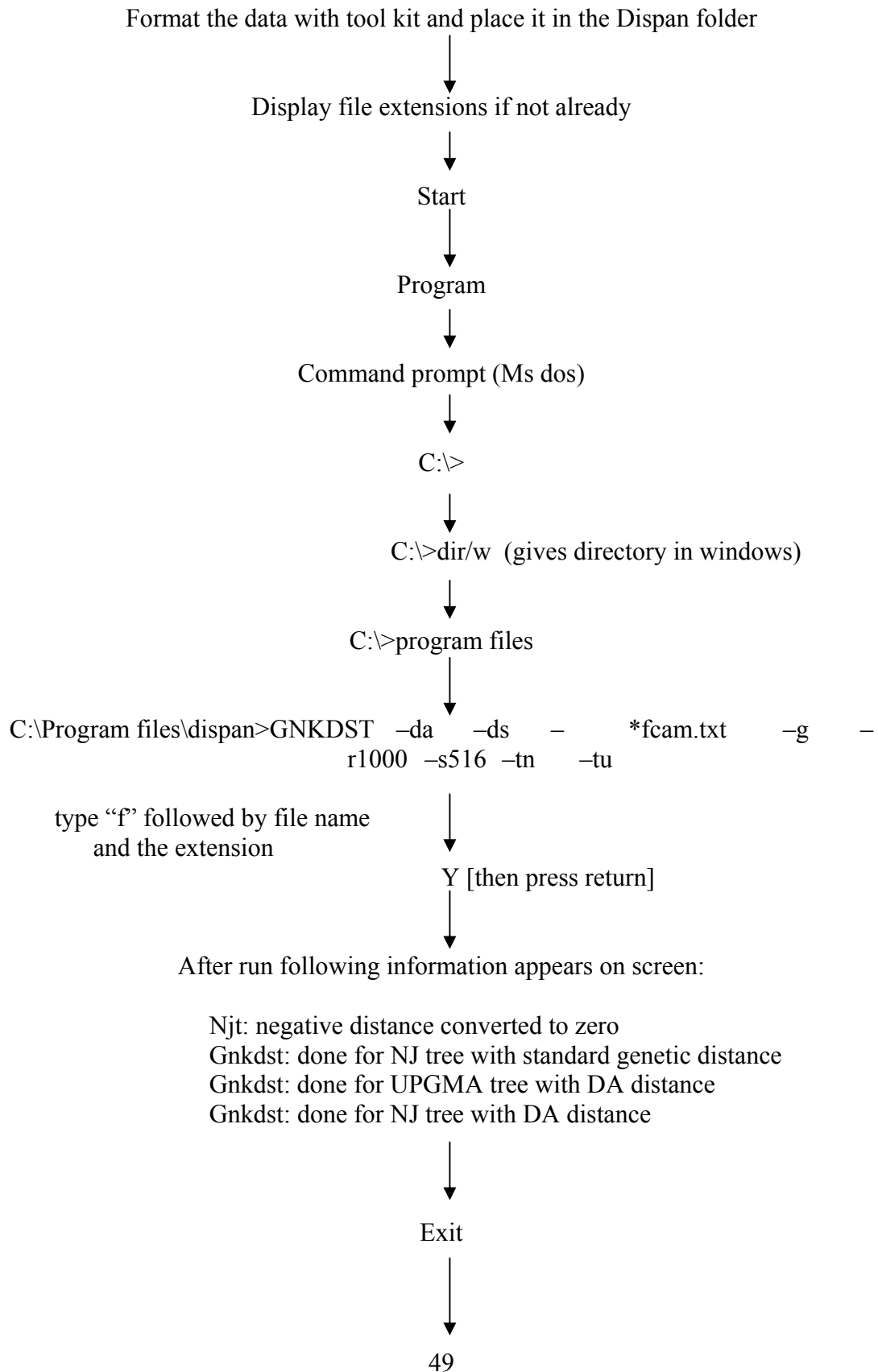




- Save
- a) Give appropriate name
 - b) Save as type text only
 - c) Save in dispan folder/microsat

An icon will appear that tells you that some characters are not compatible with text Only format and as to whether you want to age the document in that format. Just click Yes.

B) Dispan





Check Dispan folder to make sure you have following document

- (a) danj.tre }
- (b) daupgma.tre }
- (c) gnkdst.dst } Rename each individual file
- (d) stdnj.tre } e.g. all.danj.tre
- (e) stdipgma.tre }



Go to Ms dos prompt and change directories until you reach dispan



C:\dispan\treeview



Use cursor arrows to open appropriate file and press return



Escape
(edit tree as required)



Draw with graphics and return



Press "Print screen SysRq" key continuously for a few seconds to copy the phylogram)



Alt+return key (minimises screen)

Start



Programs



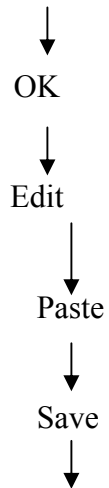
Microsoft power point



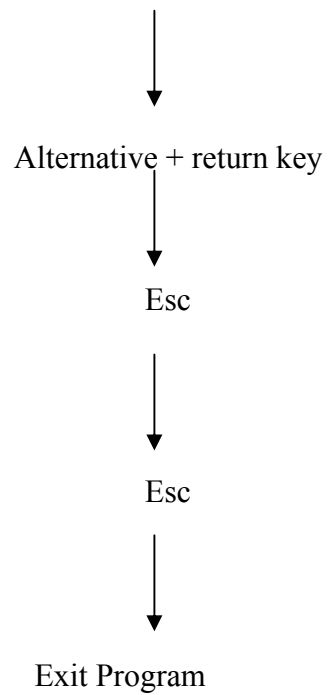
OK



Highlight blank layout



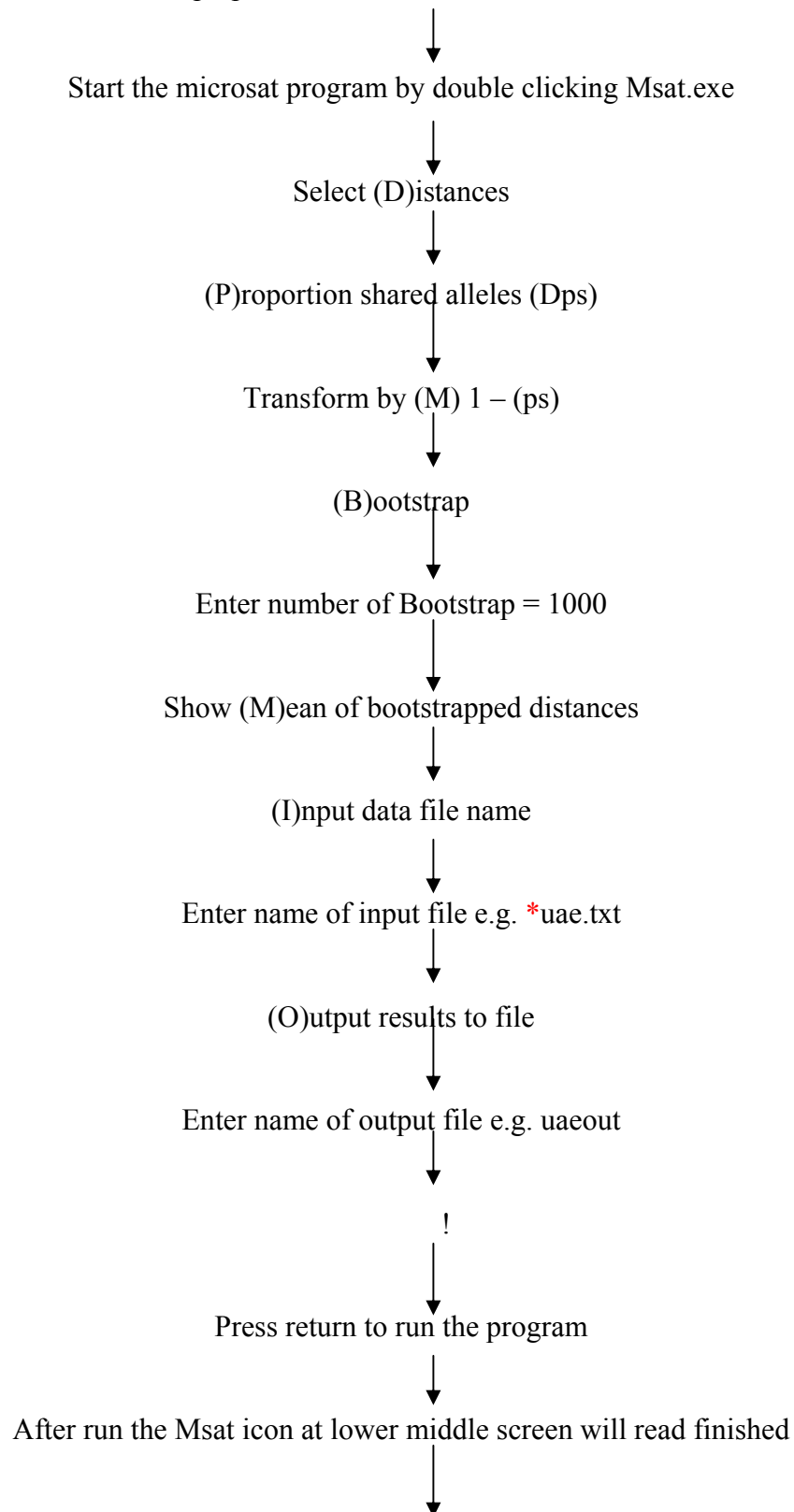
When you want to pen new file but you are in graphics
and do not want to exit Dos

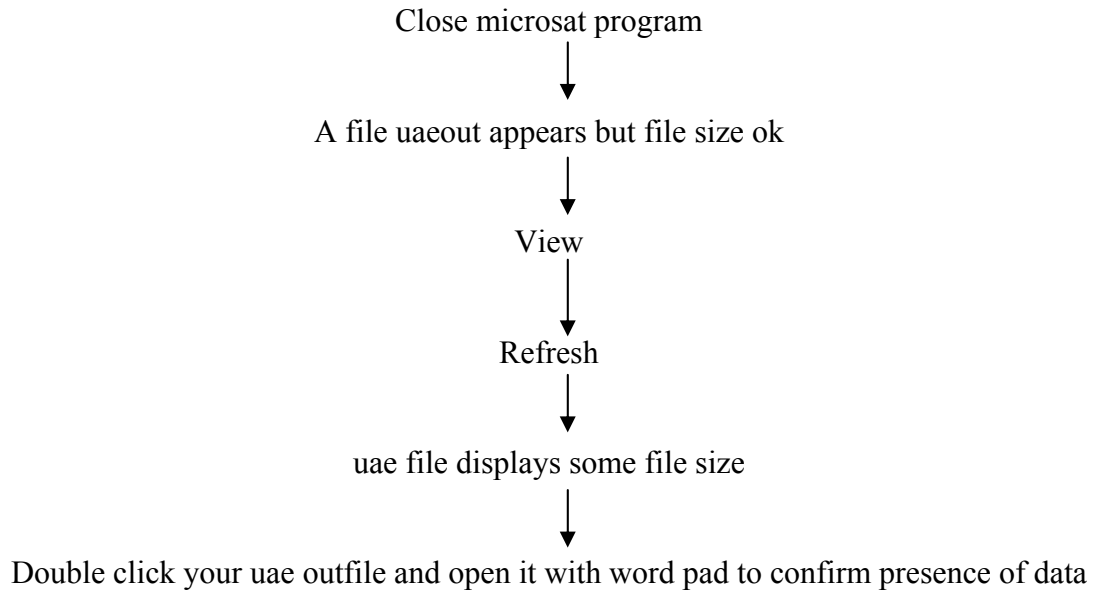


C:\program file\DISPAN>freeview

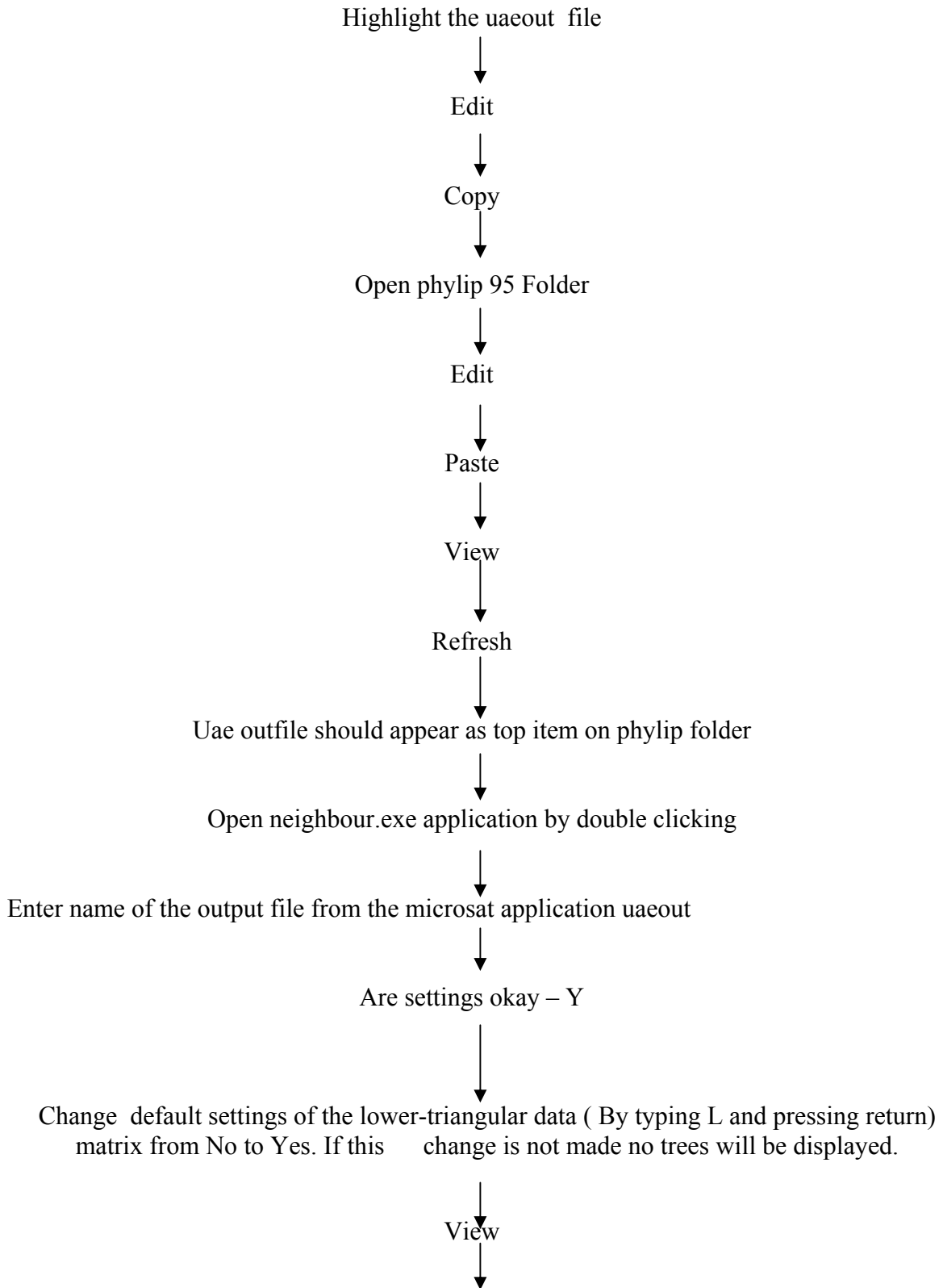
C) Microsat

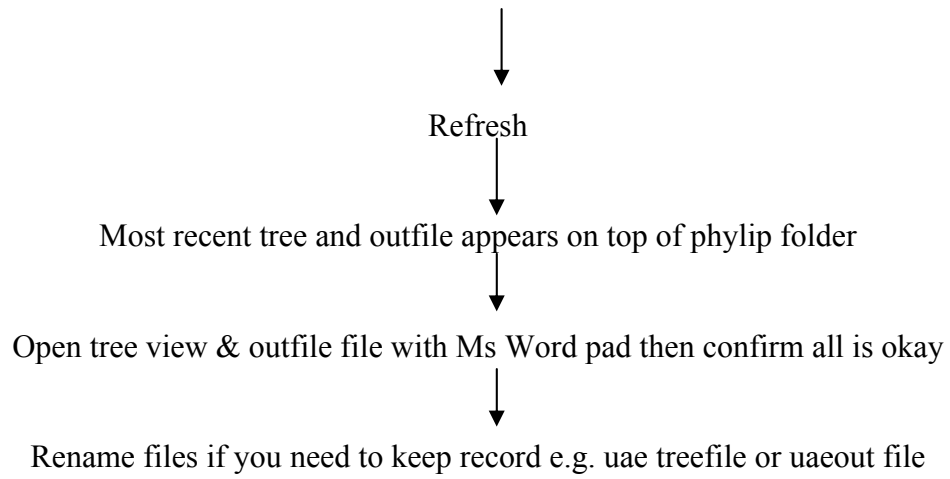
Use microsat tool kit to prepare data in microsat format- store in a Microsat folder



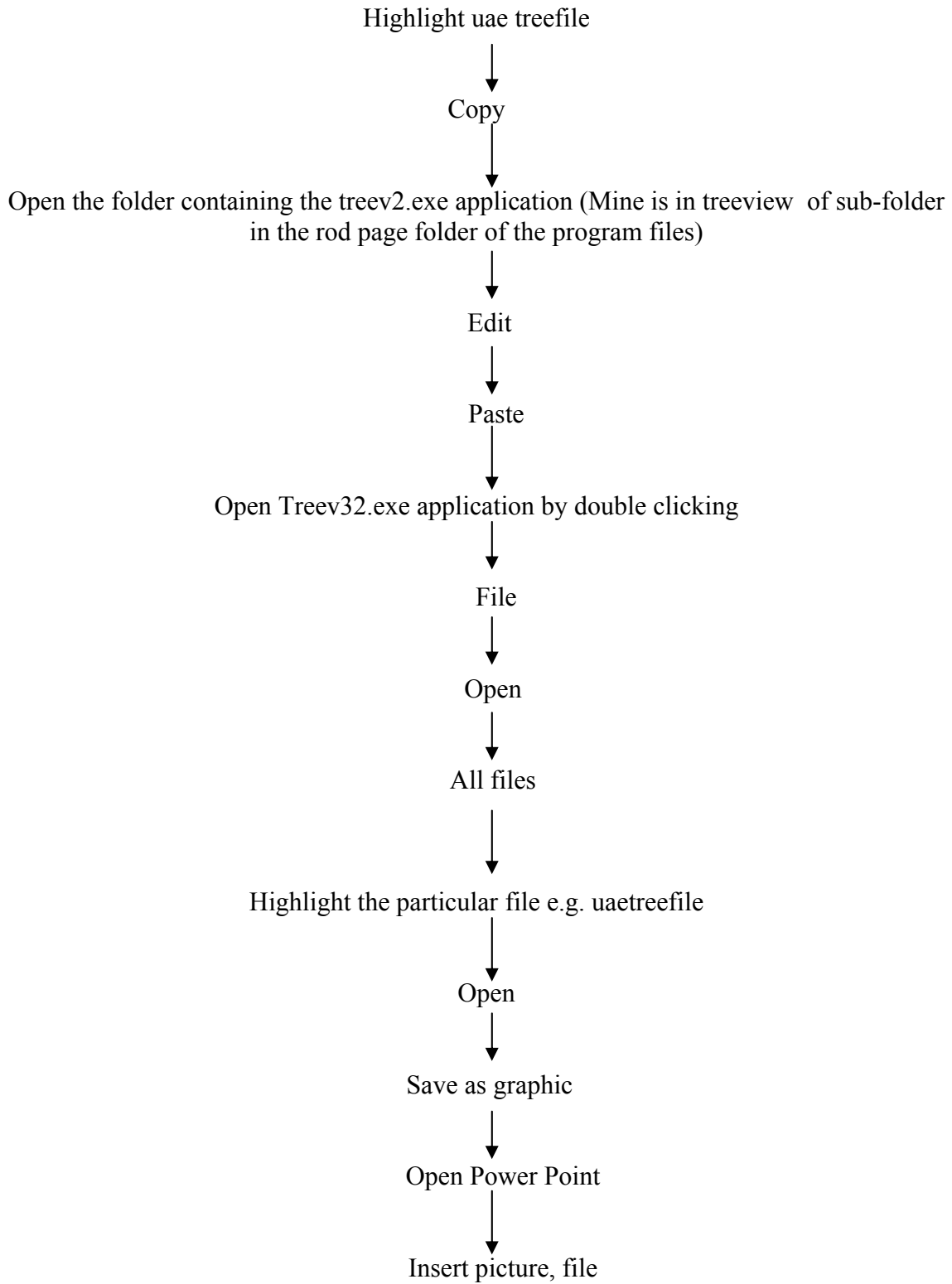


D) Phylip (neighbor.exe)

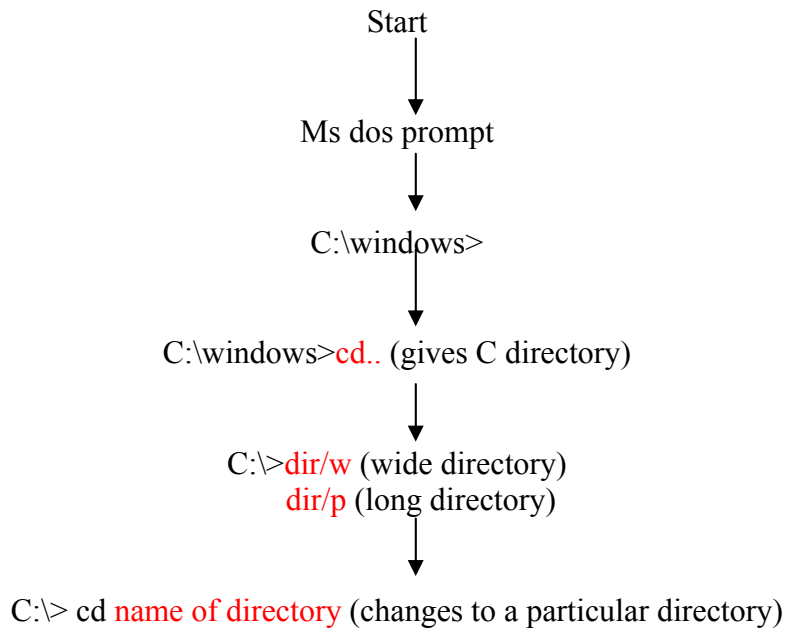




E) TreeView

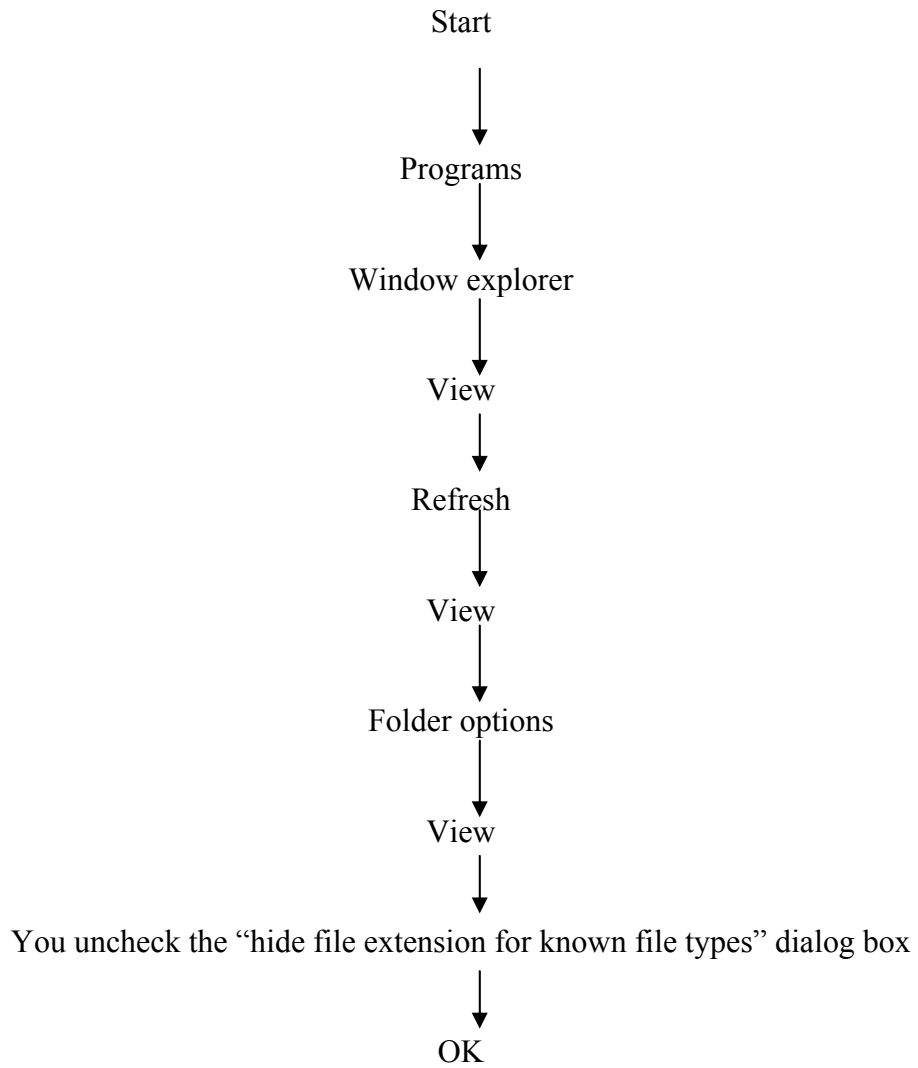


F) Ms Dos

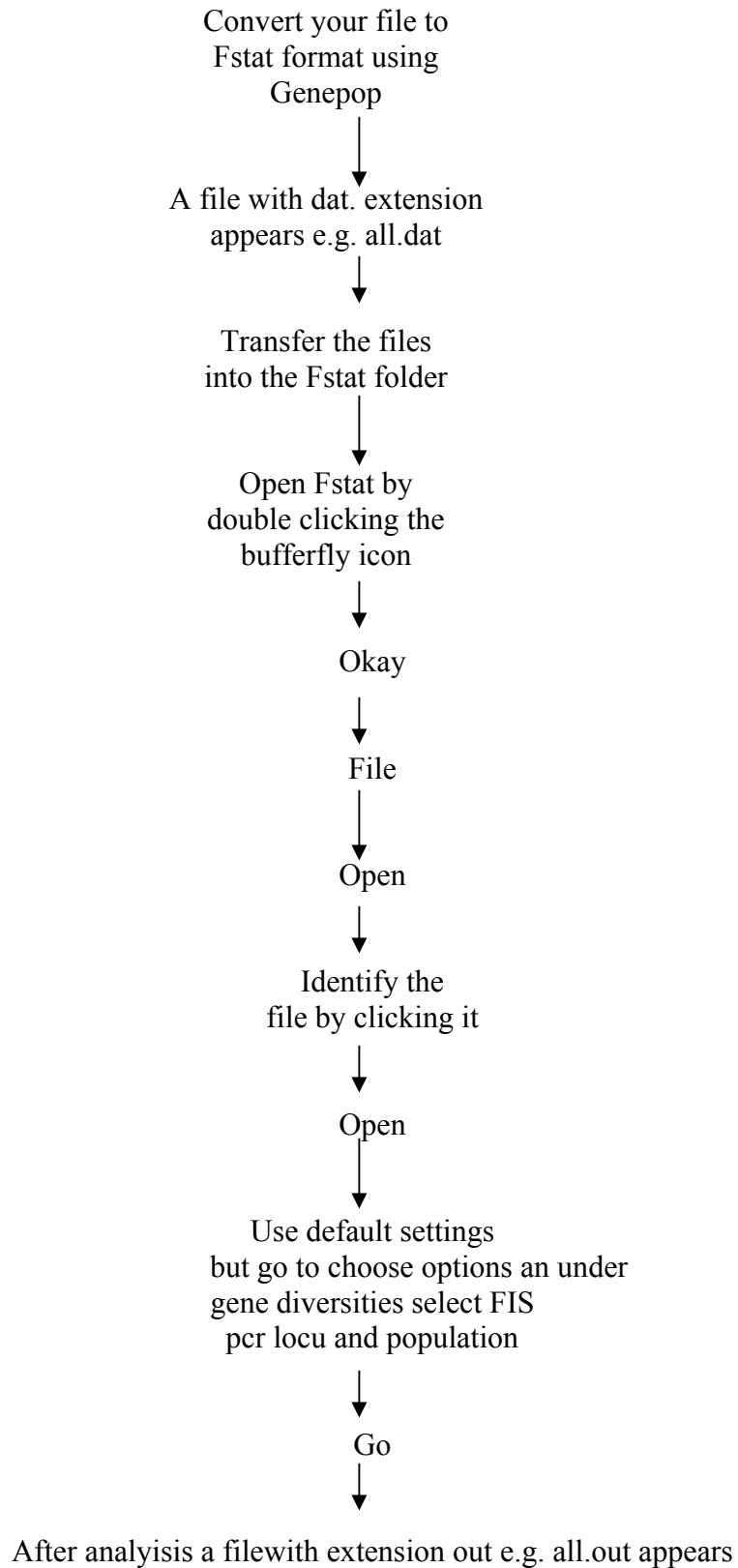


NB: cd.. (always changes to a higher root directory)

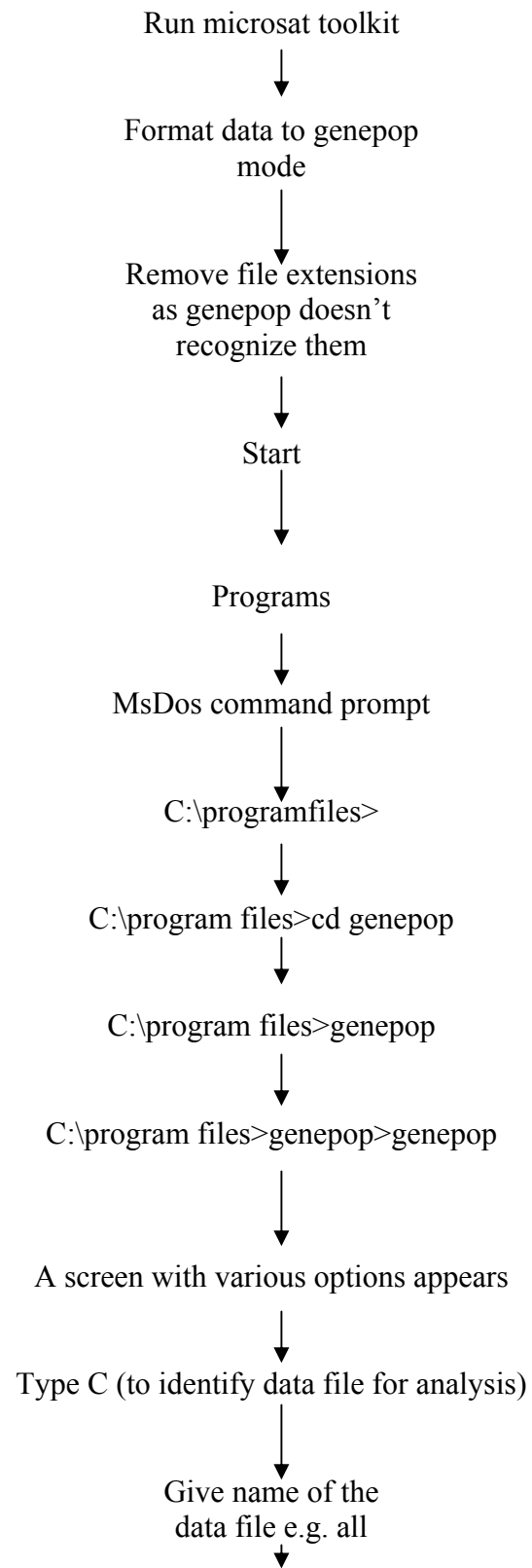
G) Displaying file extensions



H) Fstat



I) Ms Dos Genepop



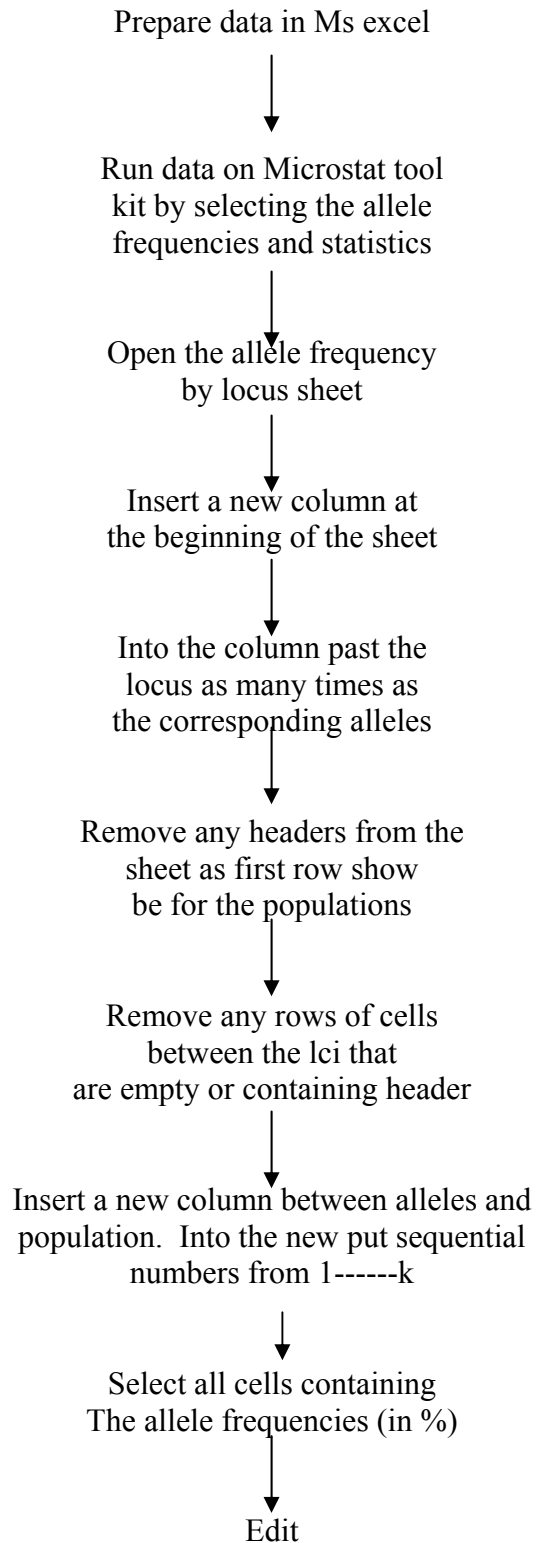
Type (N) to select
Default settings

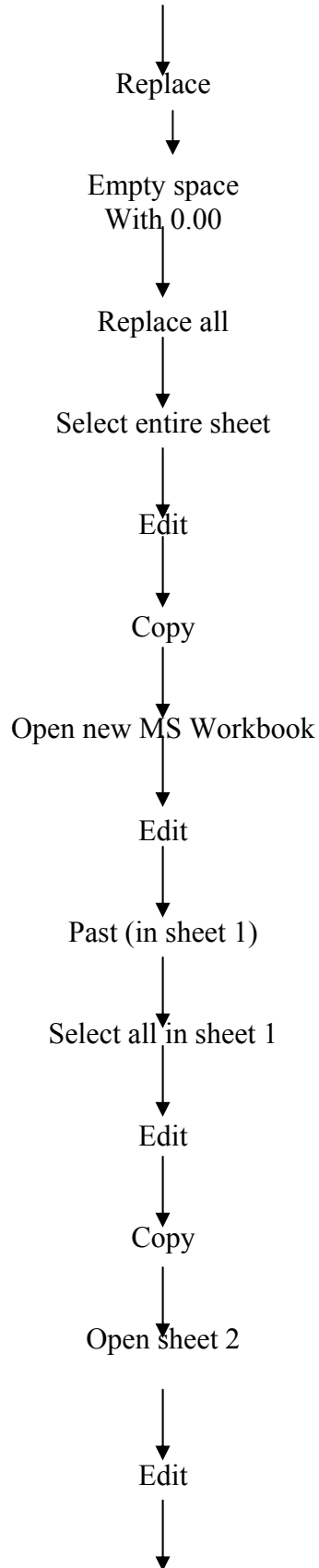


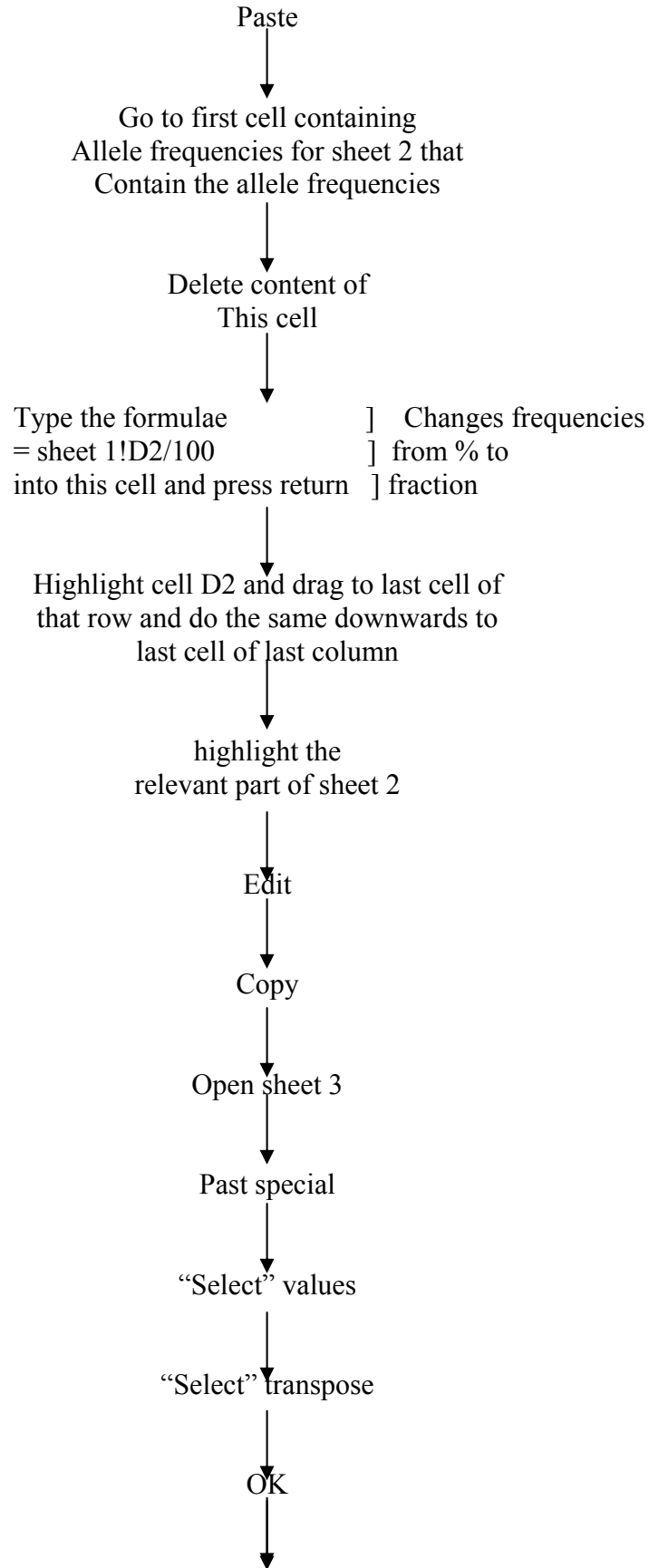
Type the necessary
numeral e.g. (1) for
Hardy-Weinbers and
run the analysis accordingly
(refer to genepop manual for details)

J) Xlstat

Data preparation









Highlight all allele frequency and convert them to three decimal places

Running Xlstat

XLSTAT



Data analysis



Correlation/PCA



Click in the input range box in your input file click the top left cell above the first population, press shift then click the last cell on right bottom that contain the allele frequencies



Click into the output range box go to the input file and in the first column, select a cell that is some say (5-10 cells) from the last population



OK



Once the plot is generated you can edit it by right clicking and choosing chart options or format plot area or chart area



You can copy & paste it in Word



When you have pasted in Word you can

↓

also edit the figure by choosing chart object
and then edit. You can now edit the figure as above.

Over 250 data entries

If you are dealing with over 250 entries, you will have to remove some that are contributing very little to the three component Axis. So go to the vector table disregard the negative sign. Note all labels with the least percentage contribution, remove them. Go to original table and replace them with our new data. Alternatively, if you have a large number of individuals in a given population, you may consider reducing the size.

Interpretation of data

Eigen value % = gives an indication as to what component as to what components contributed to what %

Co-ordinate of observation on principal axis.

- 1) Insert column of cells between the first three axis
- 2) Insert populations into the blank cells of the column
- 3) Sort by descending

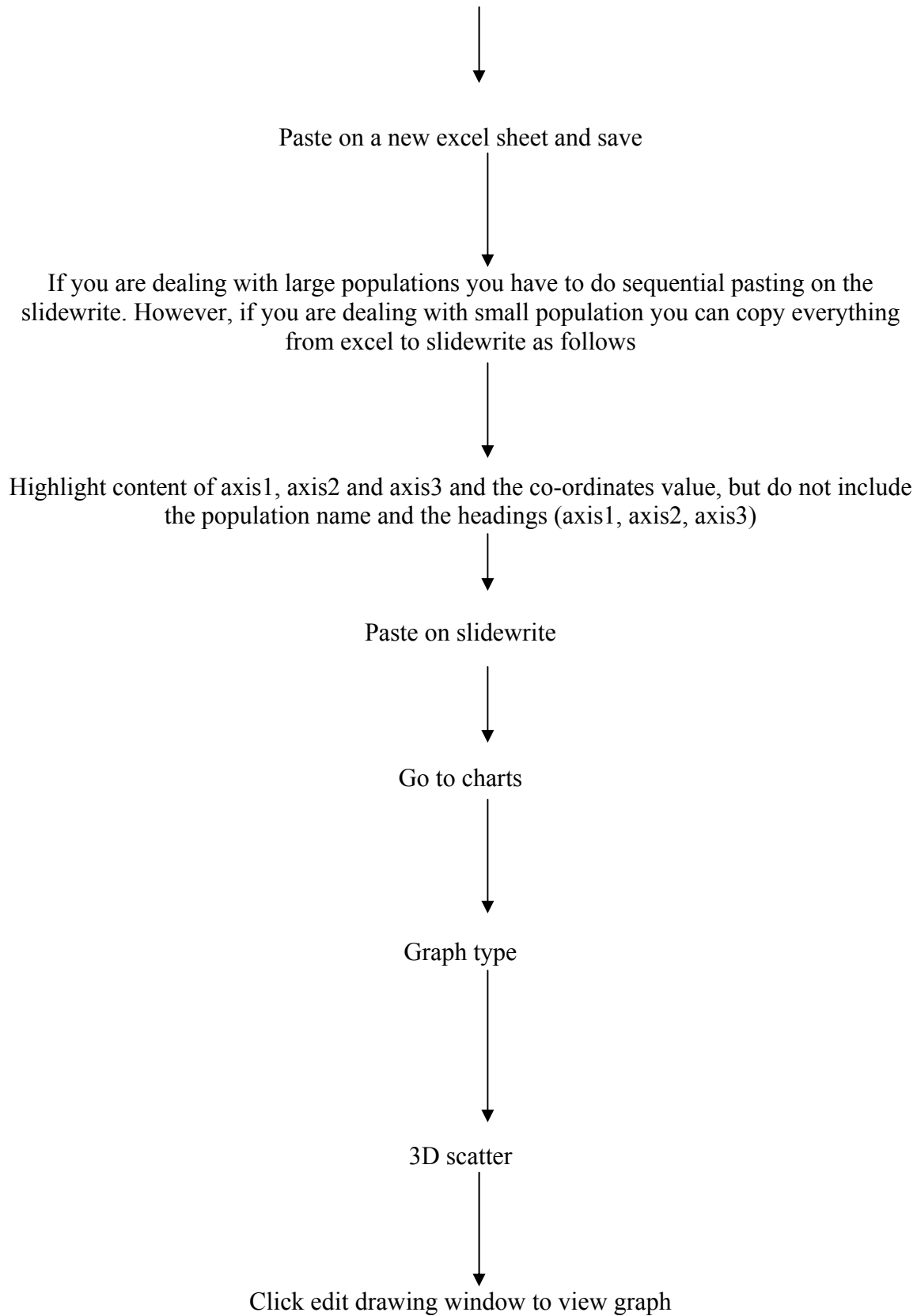
Constructing a 3D figure of PCA using slidewrite 4.0

Go to the PCA table containing “coordinates of observation on principal axis

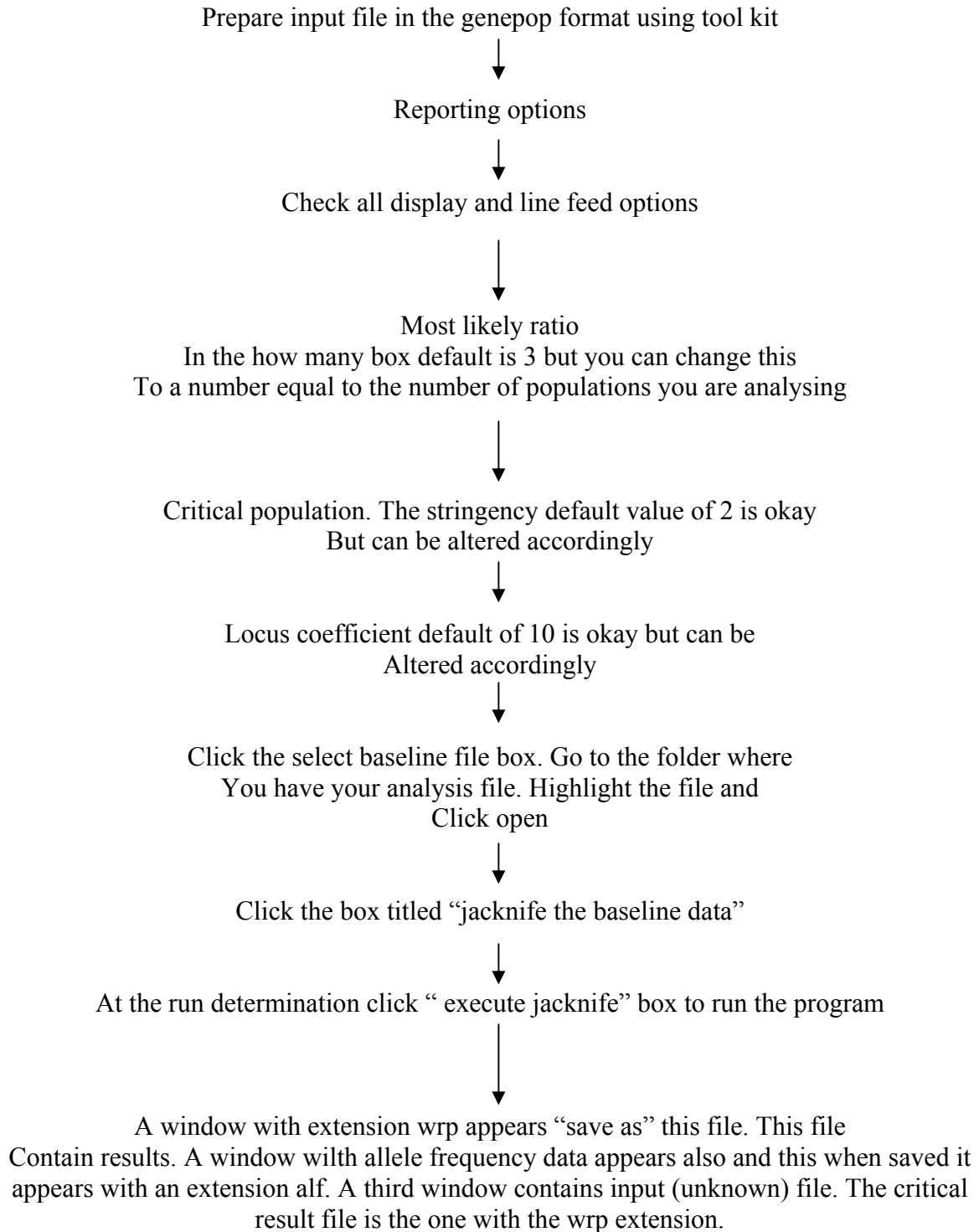
↓

Select and copy the table but limit only to population and the first three axis

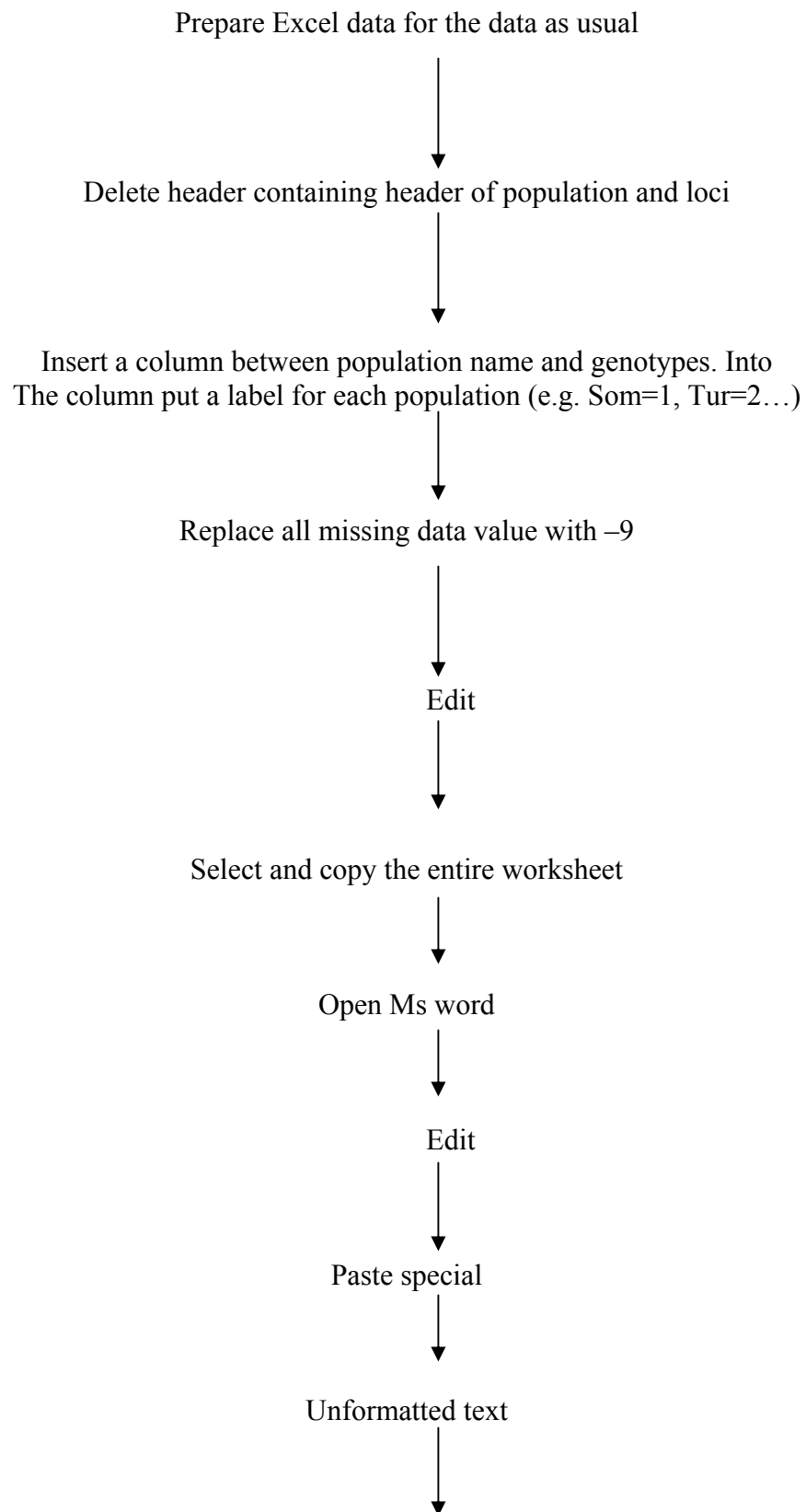
↓

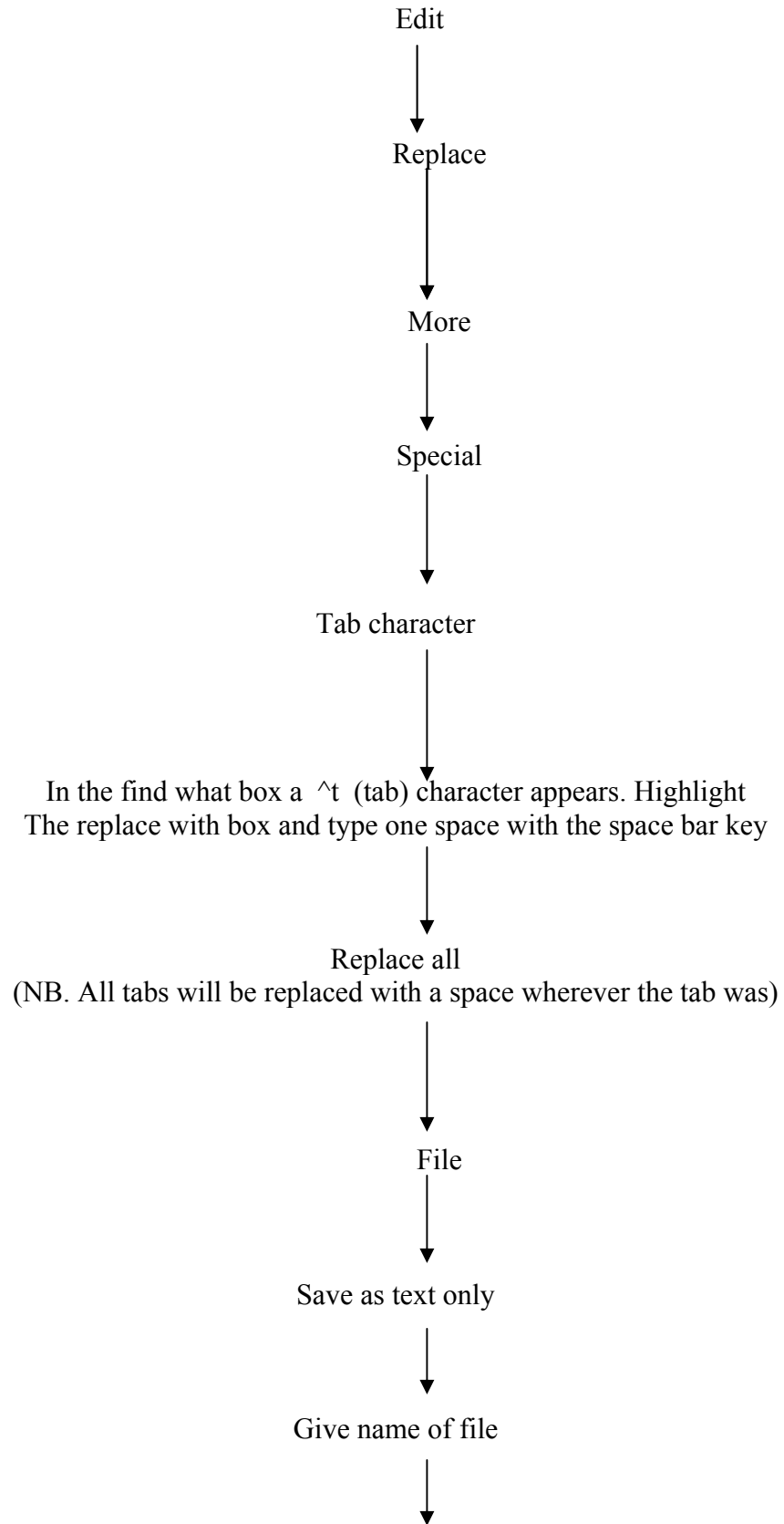


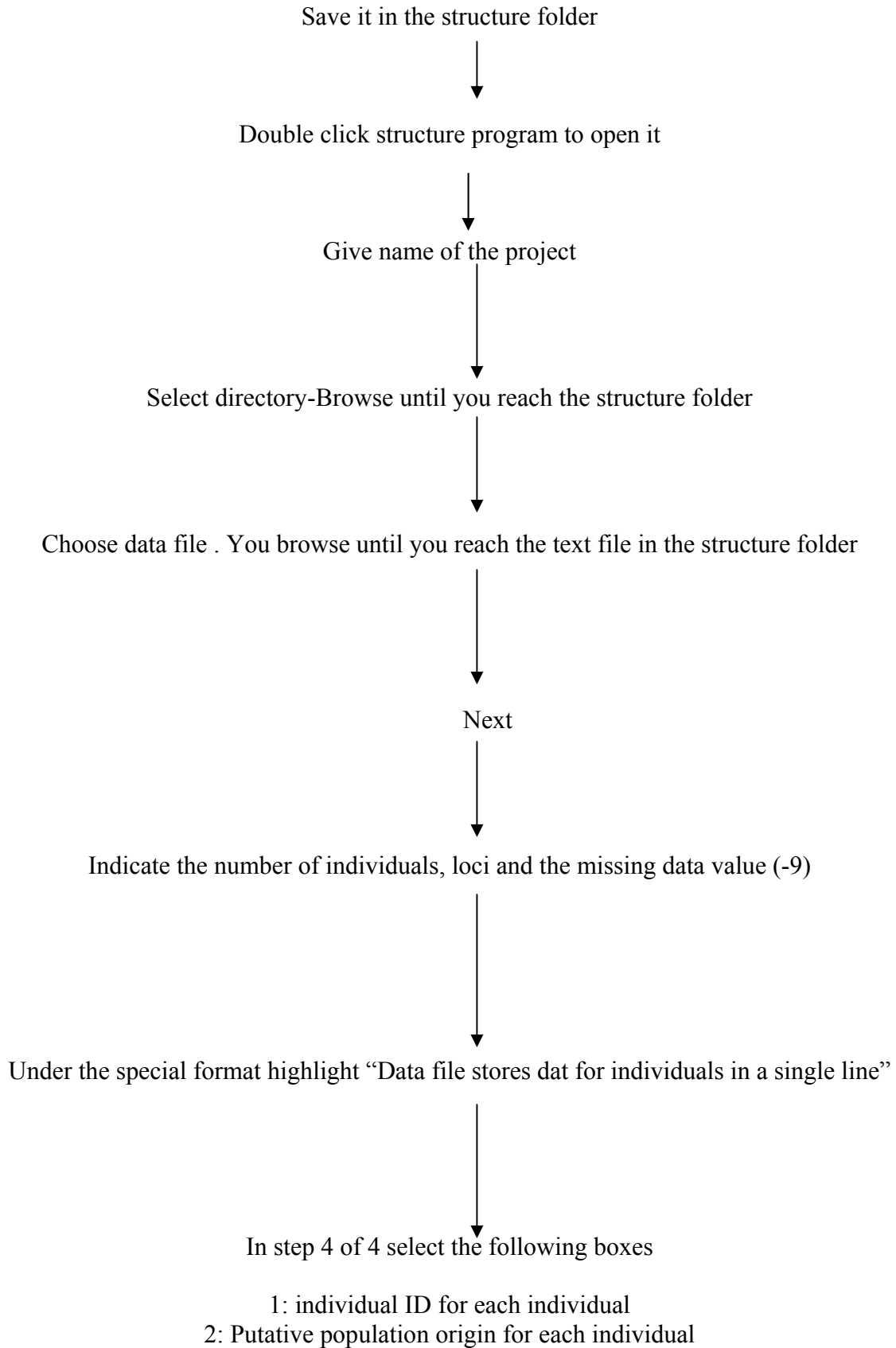
K) WhichRun



L) Structure







↓
Finish

↓
Proceed

↓
If successful the program will open an excel like project data

↓
Parameter se

↓
New

↓
Length of burning (10,000 minimum-I use 50-100,000)

↓
Number of MCMC reps after burn
(100,000 minimum-I use 400000)

↓
Allele frequency model

↓
Select allele frequencies correlated

↓
Ancestry model

Select use of admixture model



Ok



An icon requesting you to provide the new parameter set is displayed

Type into the box information that identifies the run parameters
(e.g. Adm, 30KBUN,corr,100KMCMC)



Ok



Project



Start a job



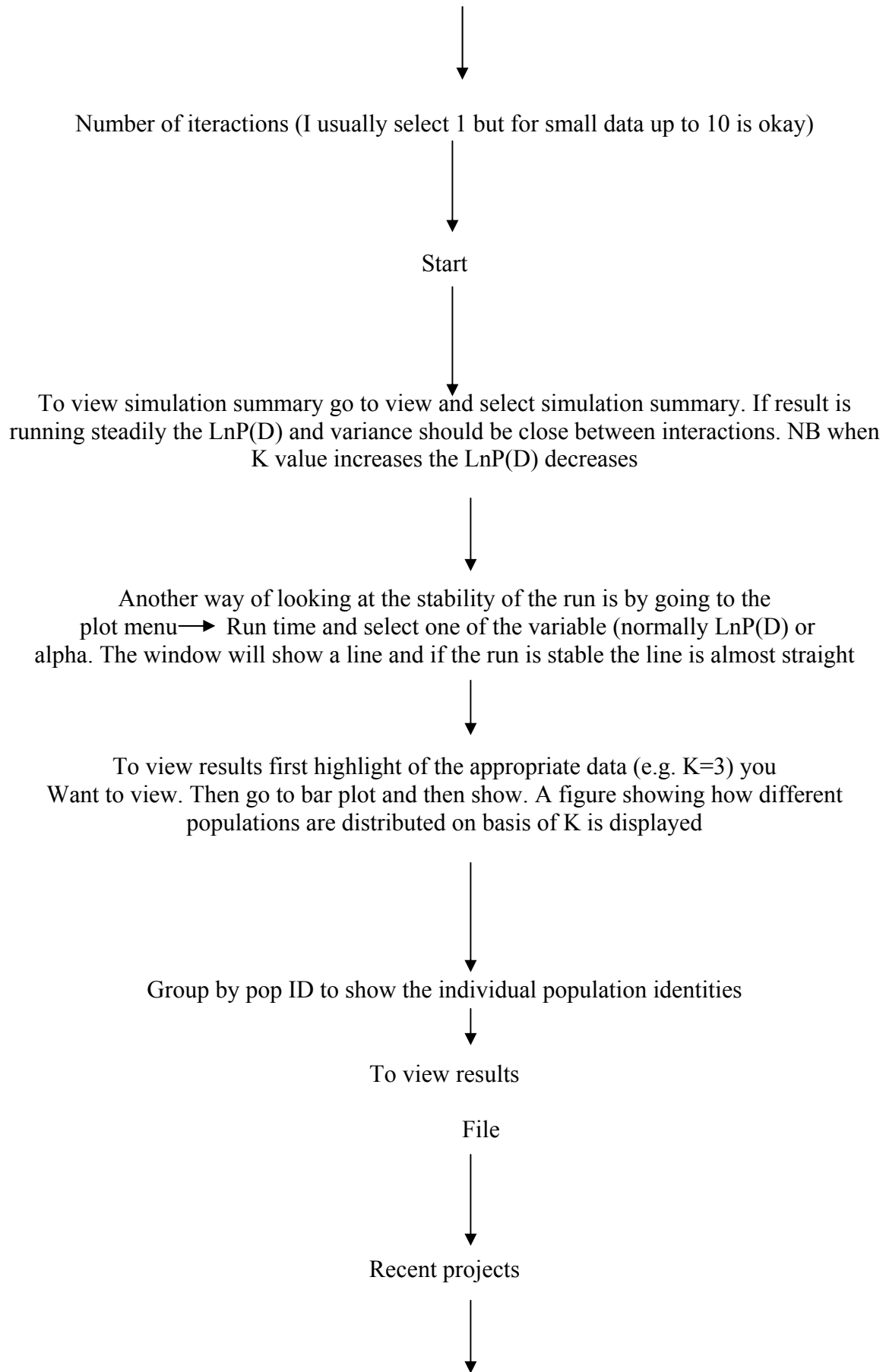
An icon is displayed allowing you to select the job to run

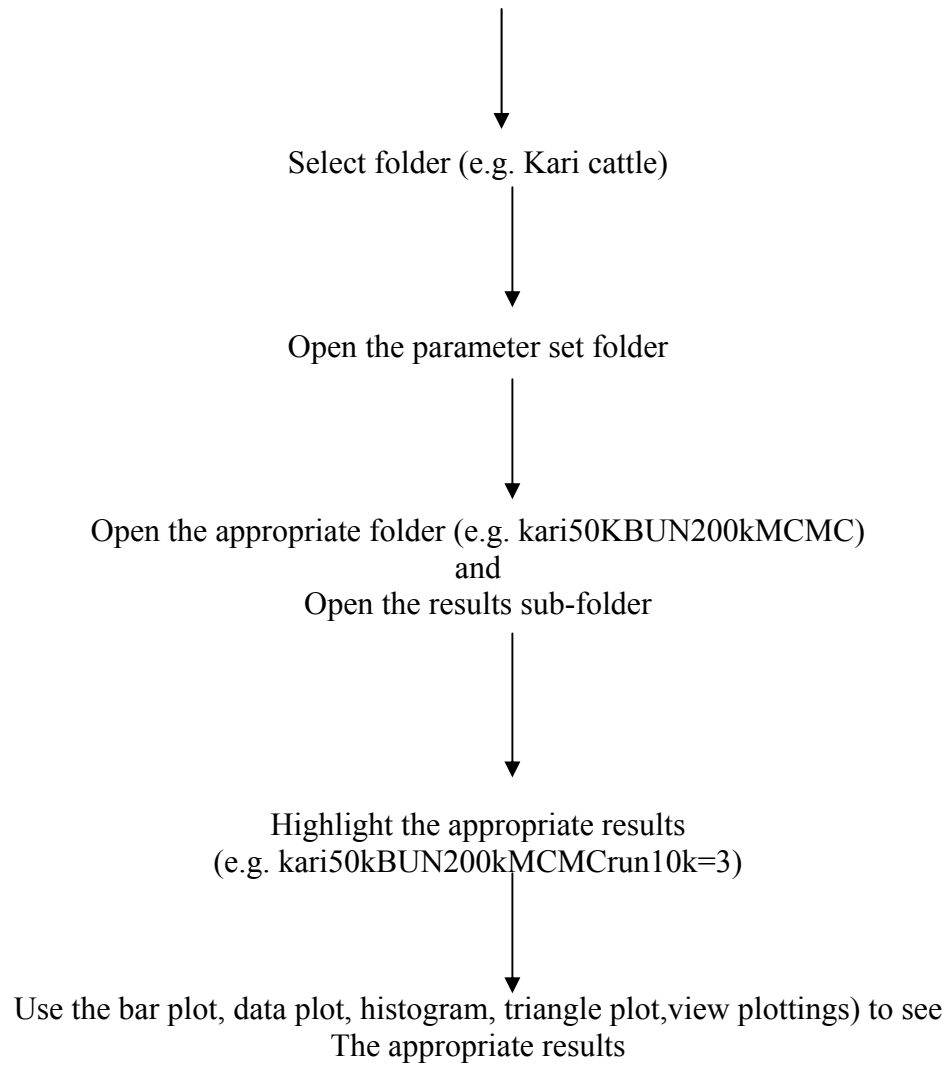


Highlight the job to run

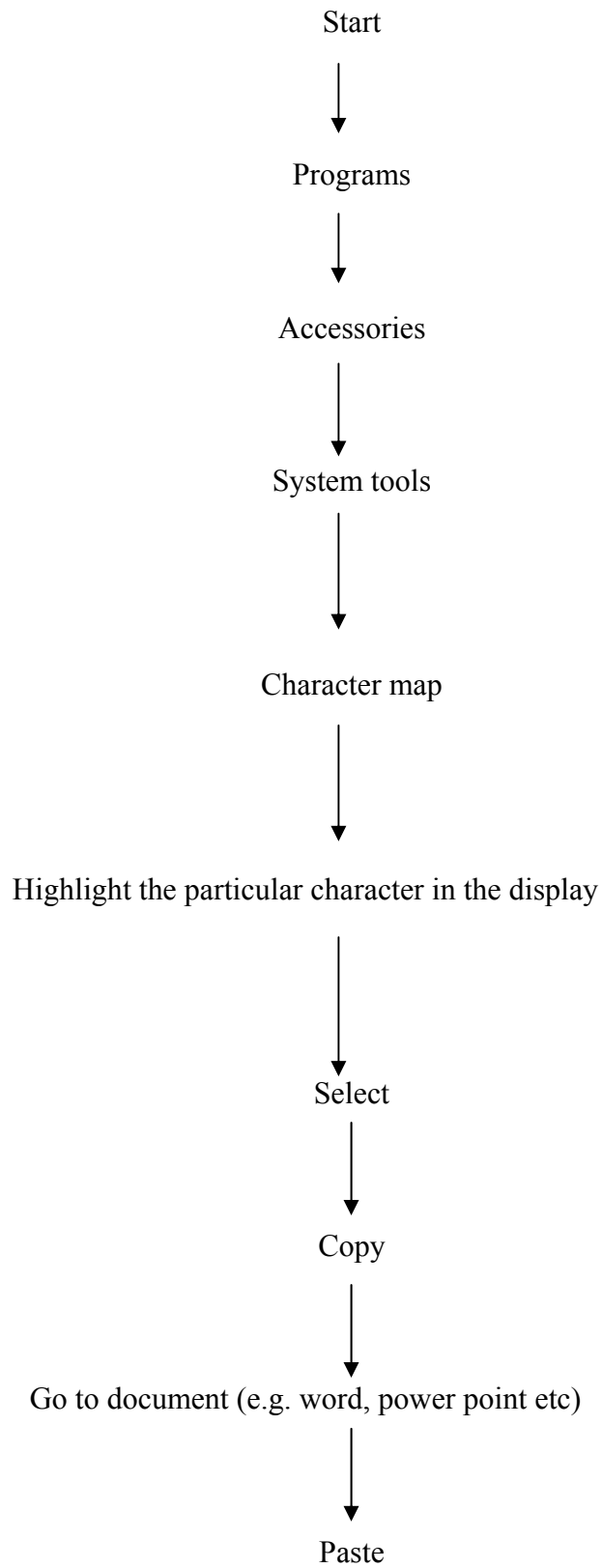


Select the appropriate 2K” from 2 to a maximum equal to maximum number
Of populations if desired





M) Special characters (e.g. \acute{E} , μ)



N) Bottleneck

Input file in genepop format



For highest accuracy use two phase mutation (TPM) and wilcoxon Signed rank test. For wilcoxon signed rank test use the probability (one tail for H excess) and score at $P < 0.05$, $P < 0.01$ $p < 0.001$



Select mode-shift box to check if it is an L-shaped distribution for Stable populations



Infinite allele model (IAM-F statistics) will underestimate variance (cause high bottlenecks) and single mutation model (R-SMM) will overestimate mutation (had to see bottlenecks) and TPM is inbetween the two



Sign tests will always underestimate calculations so high bottlenecks. So use TPM



For populations with over 20 microsatellite markers the standardized difference test can be used

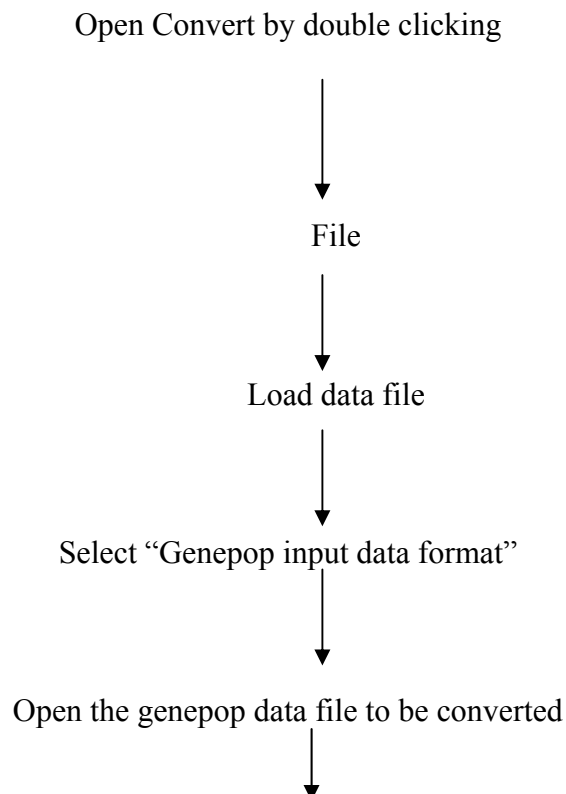
O) Convert

The program CONVERT facilitates the conversion of diploid genotypic data files into formats that can be directly read by a number of commonly used population genetic computer programs: GDA, GENEPOP, ARLEQUIN, POPGENE, MICROSAT, PHYLIP and STRUCTURE.

The input file for CONVERT follows a 'standard' format that can be easily obtained via an EXCEL file containing the genotypic data. CONVERT can also read in input data files in GENEPOP format. Note that CONVERT can only read data files in these two formats (CONVERT format or GENEPOP format).

Since we prepare all our microsat data in a standard excel format. I prefer to run this format in tool kit and convert it to genepop 2 digit format using genepop. This genepop is the input format that I use on CONVERT for conversion to an input file that can be read by one of the above listed softwares

*This software can also be used for computing allele frequencies and identification of **private alleles**



A table showing population no, last individual and
Population name is displayed

↓
Ok

Choose the desired format for file conversion (e.g. Popgene)

↓
Save

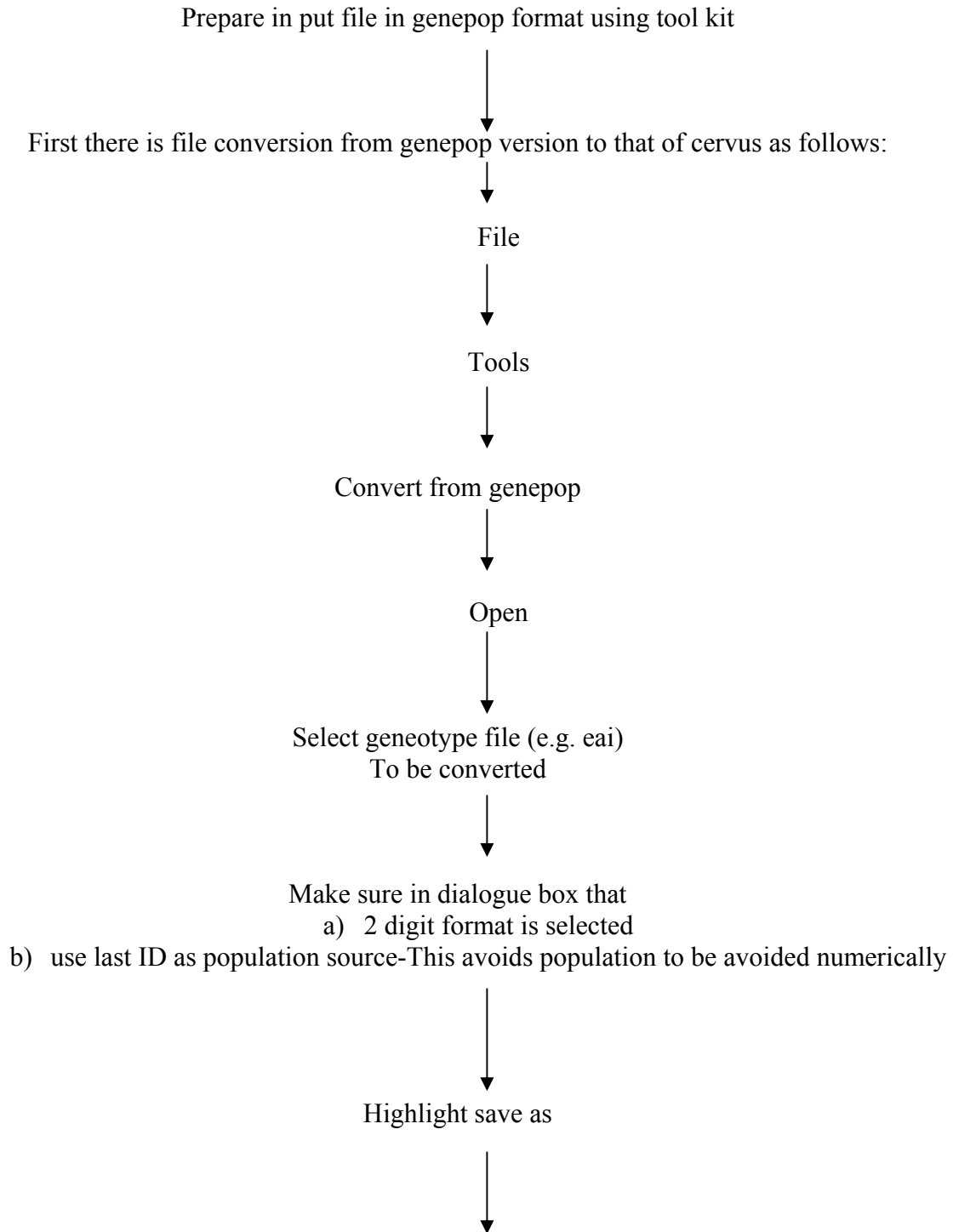
↓
The file converted to the desired format (e.g. popgene with the extension dat)
Is saved

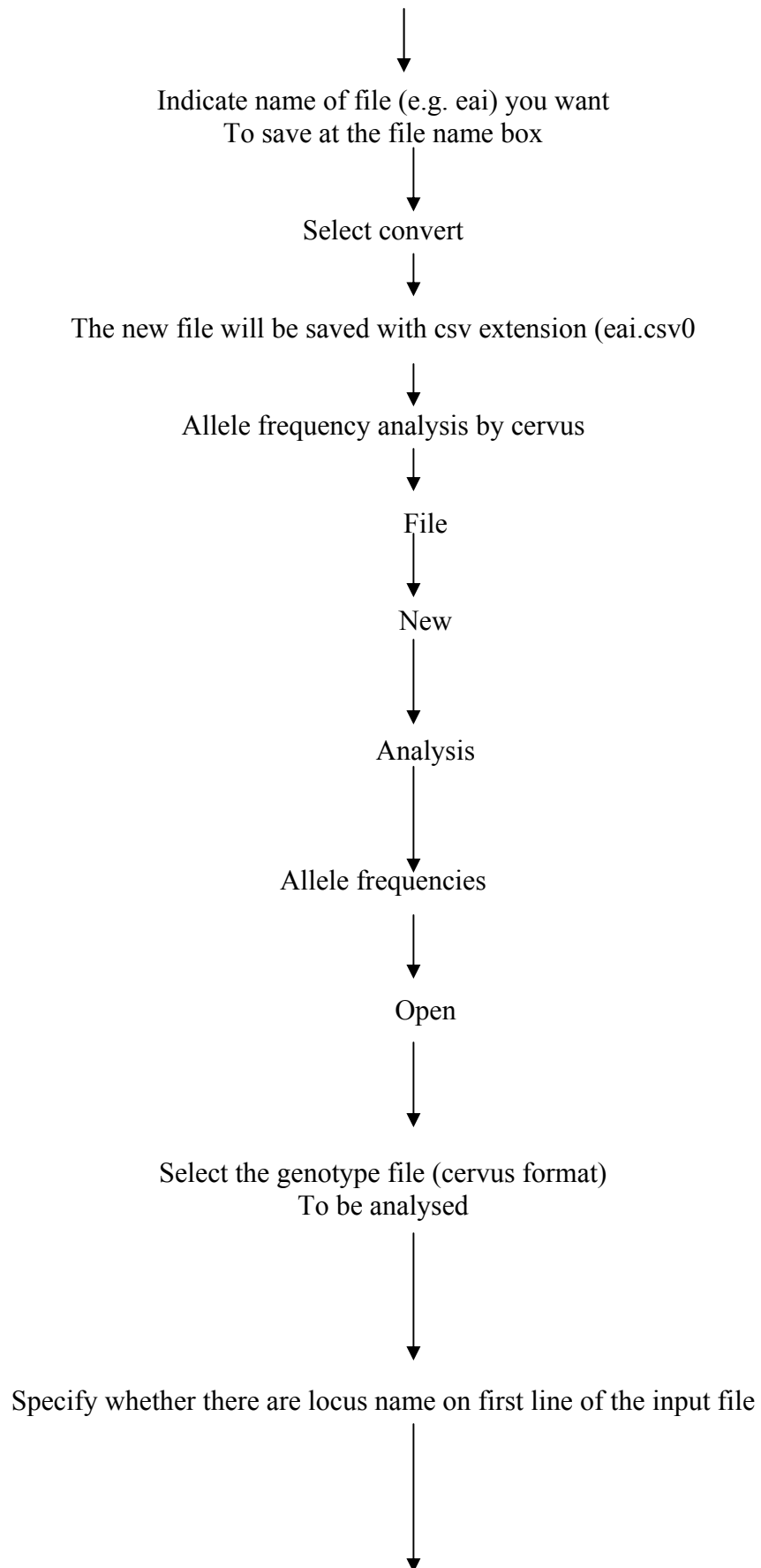
↓
A box showing successful conversion is displayed

* For computing of allele frequencies and identification of private alleles follow the flow chart but at the position marked “#” select the box “produce table of allele frequencies and continue with other steps in the flow chart.

P) Cervus

Provides single locus statistics for each microsatellite marker





Select which column in the genotype file contains the first allele of
Of the first locus



Specify the total number of loci present in the input file



Save as



Choose name of the output file (e.g. eat)
An extension txt (eat.txt) will appear



Save



Select boxes with H_W test, Yates correction,
Freq. Estimates appropriately as needed