Renate Eberl
Ecl 290

Microsatellite Toolkit and CONVERT

To start:
In My documents create a file “ECL290_work”
Go to the class folder – pick up- week3-5 and copy the following files to “ECL290_work” folder

data_for_MS_ER.xls
Modified_data_for_MS_ER.xls
GenePop3D_ER.txt
Data_for_convert_ER.txt
Modified_data_for_convert_ER.txt

this is necessary because you can not save files on the class folder

Summary of file formats in the two separate programs

	convert	ms toolkit
Arlequin	x	x
Dispan		x
Fstat		x
GDA	x	
Genepop *	x	x
Microsat	x	x
Phylip	x	
PopGene	x	
Structure	x	

*Genetix, Bottleneck, WhichRun, GeneClass

I) Microsatellite toolkit http://animalgenomics.ucd.ie/sdepark/ms-toolkit/

1) Installation: *** this instructions are for installing it on your personal computer*
1) Put the file MS_tools.xla and associated help file MS_tools.hlp into the folder where you want it to remain
2) Open Excel and select the 'Tools' menu, then 'Add-ins'.
3) Click 'Browse', then navigate to MS_tools.xla, select it and click 'OK'. 4) To load the add-in, select 'Add-ins' from the 'Tools' menu, and check the box next to 'Microsatellite tools'.
you may only need to do this the first time you use it

2) open the file “data_for_MS_ER” from the Ecl290_work folder
*******if the MS_tookkit is already added then it should be on the pulldown tools menu, if not you need to install it*
→ All programs → class software → Ecl290 → MS_tools.xla
A security warning will pop up → enable macros

3) Switching two and one column format:
tools → microsatellites →diploid data conversion → from two to one column format
a new worksheet (1ColData) in one column format will be created, check out this format

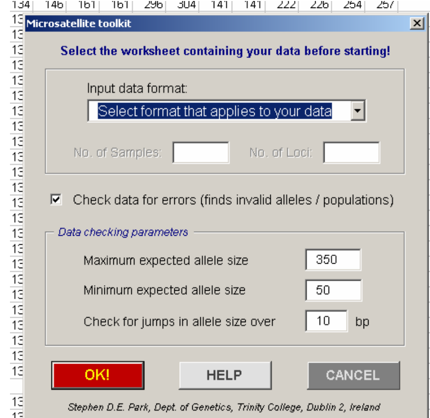
delete this worksheet and return to original data worksheet (Sheet 1)

<u>4) Check the data set:</u>

***This Procedure allows to find errors or inconsistencies in your dataset*
***Alleles are checked and classed as invalid if they are:*

- *Non-numeric (?)*
- *Non-integer*
- *Negative*
- *Below a specified lower threshold level*

in Sheet1→**tools → microsatellites → micorsatellite toolkit**



→select (data format) "diploid two column format"
select "Check data for errors" (leave data checking parameters unchanged)**→ ok**
the program will create a worksheet 1ColData ***same as we did before***
The program warns you about suspicious data (jumps in data) and asks if you want them

double column format

	Aox27		Afu63		Afu68b		Afu56		Afu160	
CAN_1	130	130	139	139	182	182	266	274	134	146
CAN_2	130	130	135	139	166	166	274	274	134	134
WILK_1	130	130	127	139	174	178	266	266	134	146
WILK_2	130	138	139	139	182	190	266	266	?	?

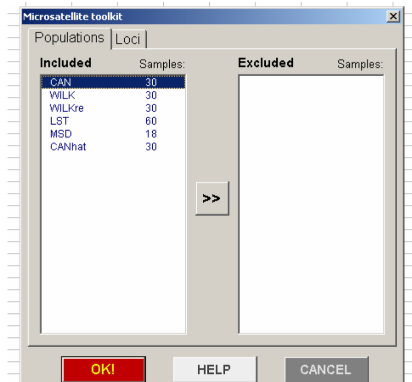
single column format

	Aox27	Afu63	Afu68b	Afu56	Afu160
CAN_1	130	139	182	266	134
	130	139	182	274	146
CAN_2	130	135	166	274	134
	130	139	166	274	134
WILK_1	130	127	174	266	134
	130	139	178	266	146
WILK_2	130	139	182	266	?
	138	139	190	266	?

highlighted click **yes "?" in program are not recognized** Edit –**find "?" and replace with "0"**
In options select "match entire cell content"

<u>5) Population and locus selection</u>

*** include and exclude populations and loci***
repeat steps in 4) until you come to this popup window
---notice! you should have 7 populations but only 6 are in window
???
Microsatellite toolkit does not distinguish between LST1 and LST2!!! → 1 pop LST n = 60
Cancel
***we need to rename these 2 populations*
In Sheet1→ edit –**find** "LST1" **replace** with "LSTA"
"LST2" with "LSTB"



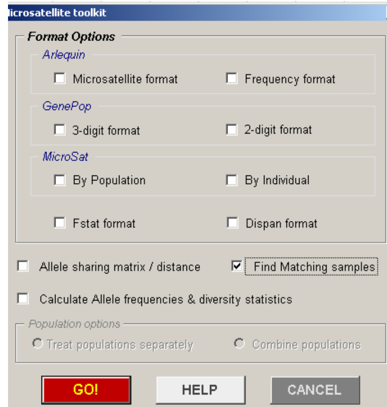
repeat above steps you should now have 7 populations → ok

<u>6) Find matching alleles:</u>

*** allows to find matching samples*

tools → microsatellites → micorsatellite toolkit →select data format and uncheck "check data for errors → ok → ok (keep all populations and loci)

****I will call this window the data conversion window***



→ select "find matching alleles"

→ enter "0" for # of non-matching alleles

The program will create a worksheet **"Matches"** showing the number of matched samples

<u>Data format conversion:</u>

Repeat steps in 6) until you get to the data conversion window again

select "GenePop 3-digit format"

make sure "treat populations separately" is selected → go

a worksheet **"GenePop3D"** is created ***format for Program Whichloci*

save worksheet "GenePop3D" as a textfile in your folder

(tools → microsatellites → save sheet as textfile) open the file and check out the format

<u>Basic Statistics:</u>

Return to excel file 1ColData worksheet

Follow steps in 6) until you get to data conversion window

select "Calculate allele frequencies and diversity statistics" and **"treat populations separately" → go …**4 new worksheets will be created

"Alleles by Pop" …lists alleles for each locus (bp and # of allele in population)

"Allele freqs" …….lists allele frequencies (%) for each locus in each population

"He and PIC" ……heterozygosity and PIC for each locus

"Stats" … # of typed loci, av.. expected (unbiased) and observed heterozygosities and average number of alleles (all with SD) for each population

II) CONVERT:

http://www.agriculture.purdue.edu/fnr/html/faculty/Rhodes/Students%20and%20Staff/glaubitz/software.htm

<u>Input file format</u>

***convert input files require a few modifications in our datafile*

Open file (data_for_MS_ER) if it is not already open

File save as "data_for_convert" in your folder save as **type 'Text(Tab delimited)(*.txt)**'.

open it in WORDPAD (or NOTEPAD).

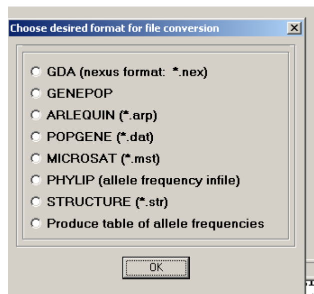
on top of page type

Andrea's data **some description of your data 1line*
npops = 7 ** *# of populations*
nloci = 10 ** *# of loci*

***in order for Convert to distinguish between populations each population must be distinguished*
type "pop = CANADIAN_LAKE" in the line above the first sample
insert appropriate population names for each new population
or
use file **"modified_data_for_convert_ER"** that has already been converted to Convert format

2) using CONVERT

open convert by clicking on the icon
file → load data file– select "convert input data file format"
***you could also select the GenePop format that we have already created*
browse to "modified_data_for_convert_ER" in your folder
the program will give you information about the source of the datafile, its input data file format Data description,# of populations, #loci and # individuals sampled – **double check !!**



in the pop up window choose **GDA** and click **OK**
Convert creates a new file
name the output file "GDA" and to **save** to folder. **The correct file extension (*.nex) is added automatically** and the program will report that the file conversion was successful

asks if you want to do **another conversion? Yes**

***create any number of different file formats but we will look at allele frequency table instead*
select **"Produce table of allele frequencies"** save as "allelefreq" in folder
another conversion? No
open file "allelefreq" and look for any private alleles in allele frequency table

Locus	Allele#	Size	Pop1	Pop2	Pop3	Pop4	Pop5	Pop6	Pop7	Overall	Private?
Afu63	1	127	0.0000	0.2000	0.3333	0.2833	0.1852	0.1944	0.0000	0.1692	
Afu63	2	135	0.1333	0.1500	0.0667	0.0667	0.0000	0.2222	0.1000	0.1000	
Afu63	3	139	0.7500	0.5333	0.4667	0.3000	0.5000	0.5000	0.7500	0.5462	
Afu63	4	143	0.1167	0.1167	0.1333	0.3500	0.1667	0.0833	0.1500	0.1641	
Afu63	5	147	0.0000	0.0000	0.0000	0.0000	0.1481	0.0000	0.0000	0.0205	LAKE_SUP2
Afu63	# samples:		30	30	30	30	27	18	30	195	