

# Γενετική Πληθυσμών και Εξέλιξη

## 2<sup>η</sup> άσκηση

**Εισαγωγή στη χρήση λογισμικού για τον υπολογισμό και τη γραφική απεικόνιση παραμέτρων ανισορροπίας σύνδεσης και την επιλογή tagging SNPs**

1. Login

**Username:** mbgguest

**Password:** mbg2015guest

2. Δημιουργήστε στην Επιφάνεια Εργασίας τον φάκελο με όνομα *Haploview\_exercise*.

3. Μέσα στον φάκελο *Haploview\_exercise*, που δημιουργήσατε στο Βήμα 2, δημιουργήστε ένα φάκελο με όνομα *Data* και ένα φάκελο με όνομα *results*.

4. Μέσα στον φάκελο *results* που δημιουργήσατε στο Βήμα 3, δημιουργήστε ένα φάκελο με όνομα *frqs* και ένα φάκελο με όνομα *lds*.

5. Πηγαίετε στη διεύθυνση <http://www.broad.mit.edu/mpg/haploview/>

6. Εδώ παρουσιάζεται το πρόγραμμα **HAPLOVIEW**.

7. Σε τι χρησιμεύει το HAPLOVIEW;

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

8. Τι είναι απλότυπος;

.....

.....

.....

.....

9. Τι είναι ανισορροπία σύνδεσης;

.....  
.....  
.....  
.....

10. Ποιος είναι ο βασικός παράγοντας που μειώνει την ανισορροπία σύνδεσης;

.....  
.....

11. Ποιες είναι οι βασικές παράμετροι που χρησιμοποιούνται για τον υπολογισμό της ανισορροπίας σύνδεσης ανάμεσα σε δύο πολυμορφισμούς (π.χ. SNPs);

.....  
.....  
.....  
.....  
.....

Τι είναι τα tagging SNPs;

.....  
.....  
.....  
.....  
.....

12. Ας επιστρέψουμε στο πρόγραμμα **HAPLOVIEW**. Το πρόγραμμα αυτό είναι διαθέσιμο δωρεάν και μπορείτε να το κατεβάσετε όλοι στον υπολογιστή σας για να αναλύσετε τα δικά σας δεδομένα. Από το μενού στα δεξιά της σελίδας επιλέγουμε **Download**. Στην ενότητα **“Download”**, επιλέξτε το **link “Haploview.jar”** προκειμένου να κατεβάσετε μια εκτελέσιμη μορφή του προγράμματος. Στο παράθυρο που θα εμφανιστεί, επιλέγετε **“Save File”**, όπως φαίνεται στην Εικόνα 1α. Στη συνέχεια ακολουθείτε τα βήματα στις Εικόνας 1β και 1γ.

You may choose to download the Haploview Windows installer (hapinstall.exe) from:

[HapInstall.exe](#)

*\*\*You may instead choose to forgo the installer and simply download the JAR file, found below.*

Install Haploview by double-clicking the installer file. The installer will create a Haploview folder in your Start Menu. To run the program, click on "Haploview.jar" file in that folder.

#### Mac / Unix / Linux

Currently there are no installers for Unix. If you have obtained the JAR file, the program can be run as follows:

```
java -jar Haploview.jar
```

Alternatively, if your system is configured properly, you should be able to start the program by double-clicking on the JAR file.

#### JAR Archive

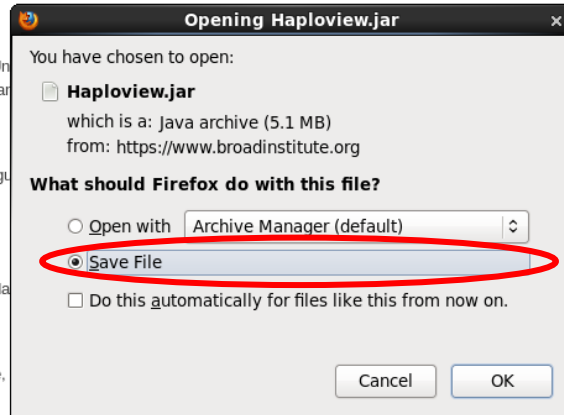
You may want to download the most recent release of the Haploview.jar file.

[Haploview.jar](#)

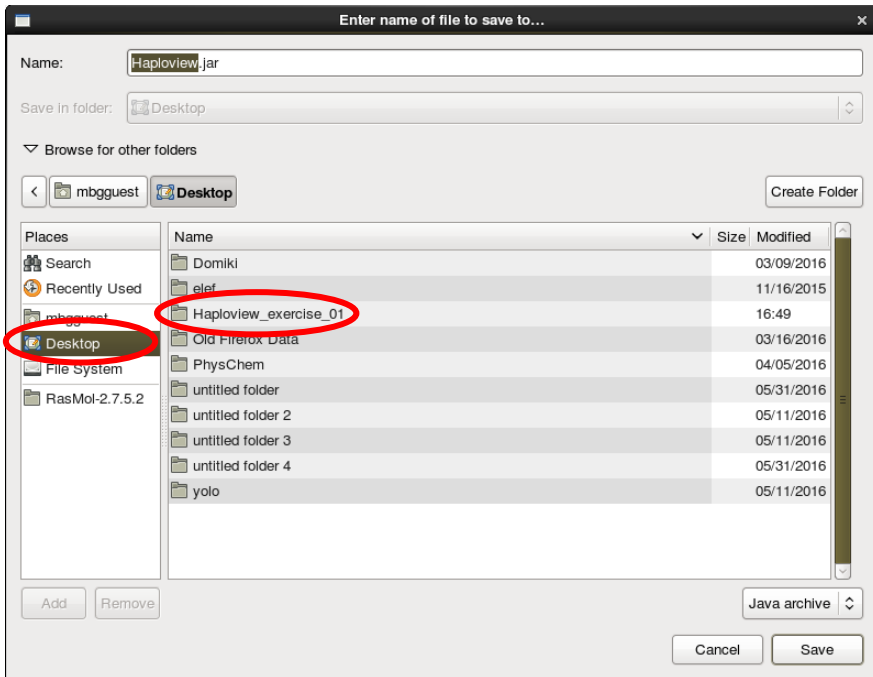
Once you have obtained the JAR file, you can run it as follows:

```
java -jar Haploview.jar
```

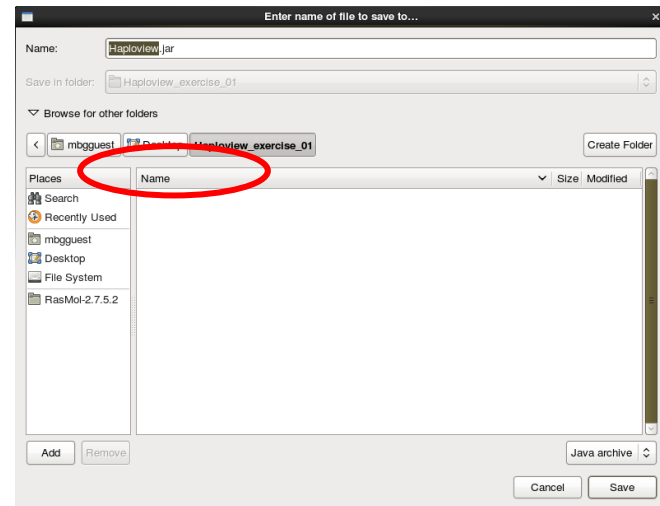
Alternatively, if your system is configured properly, you should be able to start the program by double-clicking on the JAR. You will likely want to download the [supplemental examples](#), especially if this is your first time using Haploview.



**Εικόνα 1α.** Αποθήκευση του αρχείου Haploview.jar



**Εικόνα 1β.** Αποθήκευση του αρχείου Haploview.jar



**Εικόνα 1γ.** Αποθήκευση του αρχείου Haploview.jar

13. Επιστροφή στην **ιστοσελίδα του προγράμματος HAPLOVIEW**. Επιλέγουμε από το μενού στα δεξιά, το link: *User Manual*. Τι μορφή *δεδομένων/αρχείων δέχεται* το HAPLOVIEW; Περιγράψτε τις πληροφορίες που περιέχονται σε κάθε στήλη των αρχείων με τα οποία μπορείτε να τροφοδοτήσετε το πρόγραμμα.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

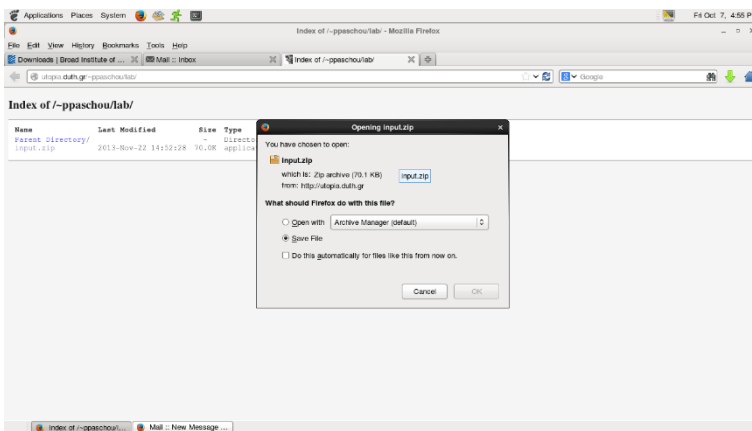
.....

.....

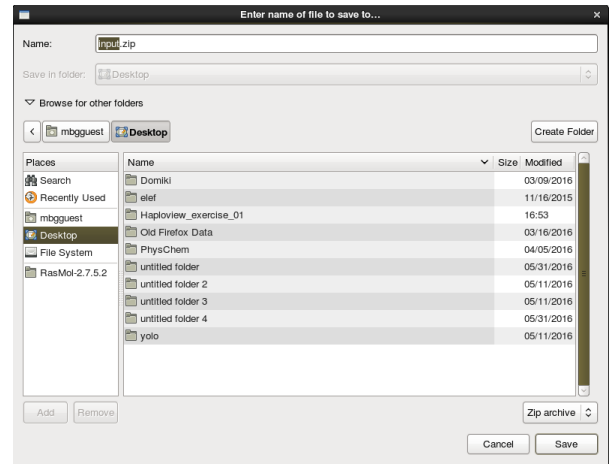
.....

GR001	GR00101	0	0	1	1	.....
GR001	GR00102	0	0	2	1	.....
GR001	GR00103	GR00101	GR00102	2	2	.....
GR002	GR00201	0	0	1	1	.....
GR002	GR00202	0	0	2	1	.....
GR002	GR00203	GR00201	GR00202	1	2	.....

14. Πηγαίνετε στην ακόλουθη ιστοσελίδα <http://utopia.duth.gr/~ppaschou/lab/> προκειμένου να κατεβάσετε τα αρχεία με τα δεδομένα των πληθυσμών που θα αναλύσετε σήμερα. Το αρχείο έχει όνομα *input.zip*. Με κλικ επάνω στο αρχείο *input.zip* αποθηκεύεται στο φάκελο *Data* (που βρίσκεται στο φάκελο *Haploview\_exercise*) (Εικόνες 2α, 2β).

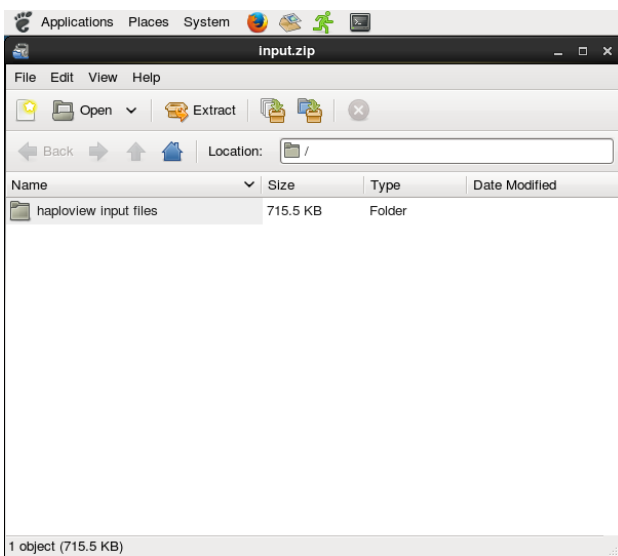


Εικόνα 2α. Αποθήκευση του αρχείου input.zip

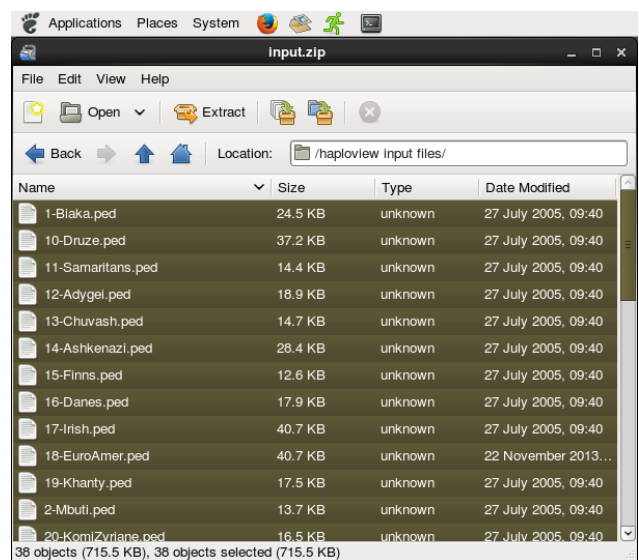


Εικόνα 2β. Αποθήκευση του αρχείου input.zip

15. Αποσυμπιέστε το αρχείο *input.zip* μέσα στο φάκελο *Data*. Διπλό κλικ στο αρχείο input.zip, οπότε εμφανίζεται η Εικόνα 3α. Στη συνέχεια, ανοίγουμε το φάκελο haploview input files (διπλό κλικ) (Εικόνα 3α) και επιλέγουμε όλα τα αρχεία με **CTRL+A**, οπότε το αποτέλεσμα θα είναι όπως της Εικόνας 3β.

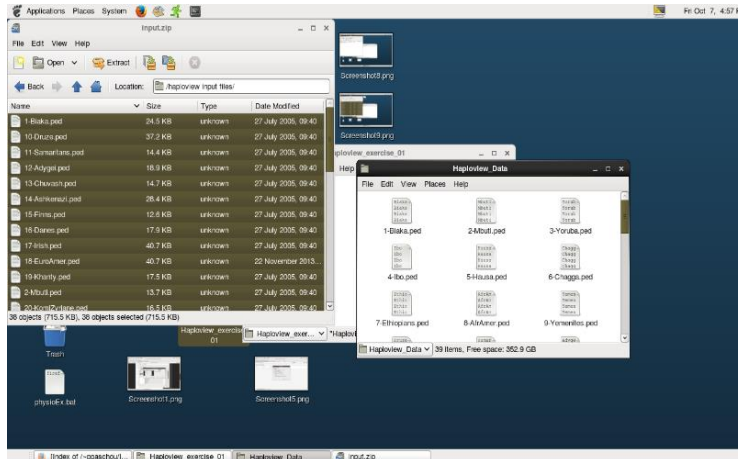


Εικόνα 3α. Αρχείο input.zip

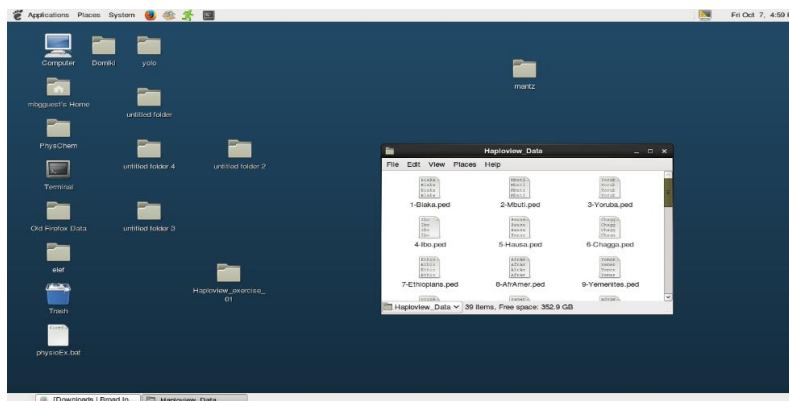


Εικόνα 3β. Επιλογή όλων των πληθυσμών.

16. Με κλικ στα επιλεγμένα αρχεία και σέρνοντάς τα μεταφέρουμε τα αρχεία στο φάκελο **Data** (Εικόνα 4α, 4β).



Εικόνα 4α. Drag and drop των αρχείων



Εικόνα 4β. Τα αρχεία δεδομένων στο φάκελο **Data**

17. Υπάρχουν 37 αρχεία με επέκταση “ped” – ένα για κάθε έναν από τους 37 πληθυσμούς. Το κάθε αρχείο έχει το όνομα του πληθυσμού. Στο τέλος του πρωτοκόλλου φαίνονται οι πληθυσμοί για τους οποίους έχετε δεδομένα. Είναι ταξινομημένοι από την Αφρική προς Ευρώπη, Ασία, Βόρεια και Νότια Αμερική, σύμφωνα με ένα πιθανό πρότυπο με το οποίο μπορεί να έγινε ο αποικισμός της γης από τον ανατομικά σύγχρονο άνθρωπο, μετά την έξοδό του από την Αφρική. Πόσες γεωγραφικές περιοχές θα μελετήσετε;

---

18. Ο κάθε πληθυσμός έχει τυποποιηθεί για 63 SNPs με τη δοκιμασία Taqman. Τα SNPs αυτά καλύπτουν μια περιοχή 900Kb περίπου στην περιοχή 17q25, κοντά στην άκρη δηλαδή του μακρού βραχίονα του χρωμοσώματος 17. Καλύπτουν το γονίδιο TBCD (ένα μοριακό συνοδό της τουμπουλίνης, απαραίτητο για τη σωστή διαμόρφωση της πρωτεΐνης στο χώρο κι επομένως τη σωστή διαμόρφωση των μικροσωληνίσκων).

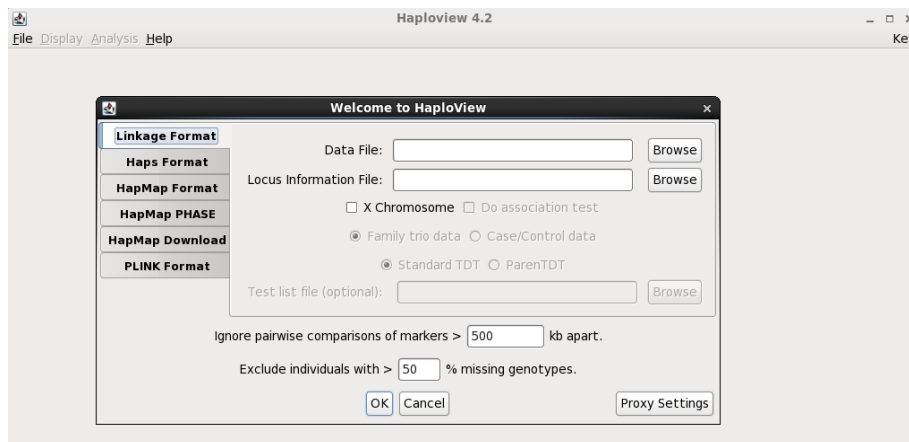
19. Ανοίξτε τα αρχεία με επέκταση “ped” χρησιμοποιώντας την εφαρμογή LibreOffice Calc. Αυτό επιτυγχάνεται με δεξί κλικ σε κάθε αρχείο. Αν δεν εμφανίζεται η συγκεκριμένη εφαρμογή στις διαθέσιμες δυνατότητες, τότε επιλέγουμε Open with Other Application... και στο νέο παράθυρο βρίσκουμε την εφαρμογή LibreOffice Calc. Για τη σωστή εμφάνιση των αποτελεσμάτων επιλέγουμε τις δυνατότητες: Merge delimiters και Space. Παρατηρήστε τη μορφή των input files. Τι μορφή έχουν;

.....  
.....  
**ΠΑΡΑΤΗΡΗΣΗ 1:** Στην περίπτωση που δεν υπάρχει το LibreOffice Calc, μπορείτε με διπλό κλικ να ανοίξετε τα αρχεία με τον κειμενογράφο

**ΠΑΡΑΤΗΡΗΣΗ 2:** Στην περίπτωση λειτουργικού συστήματος Windows και της εφαρμογής Microsoft Office, θα μπορούσατε να δείτε τα αρχεία με κατάληξη ped, χρησιμοποιώντας το Excel.

20. Θα ασχοληθούμε αρχικά με τον πληθυσμό των Adygei. Πού ζουν οι Adygei;

21. Άνοιγμα του HAPLOVIEW. Θα εμφανιστεί η ακόλουθη εικόνα.



**Εικόνα 5.** Το πρόγραμμα Haploview

22. Επιλέξτε “Linkage Format” και φορτώστε το ped file, πατώντας στο **Browse**, για τους Adygei (12-Adygei.ped) στο πεδίο Data file και το αρχείο 63snps\_info.txt, στο πεδίο Locus Information File. **Μην πατήσετε ακόμη OK.**

23. Εδώ, μέσω της επιλογής *Ignore pairwise comparisons of markers > kb apart*, έχετε επίσης τη δυνατότητα να επιλέξετε να μην γίνει έλεγχος για ανισορροπία σύνδεσης ανάμεσα σε SNPs που βρίσκονται σε μεγάλη απόσταση (π.χ. πάνω από 500kb). Τι συσχέτιση περιμένουμε για τα SNPs που βρίσκονται σε μεγάλη απόσταση και γιατί;

24. Ρυθμίστε τις παραμέτρους των συγκρίσεων ώστε να κάνουμε όλα τα κατά ζεύγη τεστ ανάμεσα στα SNPs για τα οποία έχουμε δεδομένα, ανεξάρτητα από τη μεταξύ τους απόσταση (το σχετικό πεδίο πρέπει να είναι κενό).

25. Επίσης, μπορείτε να αποκλείσετε από την ανάλυση τα άτομα που έχουν πάνω από κάποιο συγκεκριμένο ποσοστό αποτυχίας στη γονοτύπηση (*Exclude individuals with >50 % missing genotypes*). Γιατί μπορεί να θέλετε να επιλέξετε κάτι τέτοιο;

26. Αφήστε το σχετικό πεδίο κενό. Θα αναλύσετε όλα τα άτομα. Μην πατήσετε ακόμη OK.

27. Αν στα δεδομένα σας υπήρχαν ασθενείς και άτομα ελέγχου, και είχατε τυποήσει κάποια SNPs σε ένα γονίδιο με στόχο να βρείτε αν αυτό το γονίδιο εμπλέκεται στην αιτιολογία της ασθένειας, θα μπορούσατε να χρησιμοποιήσετε το HAPLOVIEW για να κάνετε έναν έλεγχο συσχέτισης (association test). Το HAPLOVIEW δέχεται εκτός από ασθενείς και άτομα ελέγχου (controls) και δεδομένα για trios (ένα παιδί με την ασθένεια που μας ενδιαφέρει και οι γονείς του). Για αυτό τον λόγο χρησιμοποιείται η μορφή linkage για τα input files, στην οποία μπορεί να δηλωθεί εκτός από το status της ασθένειας για κάθε άτομο (ασθενής/υγιής) και η σχέση των ατόμων σε μια οικογένεια.

28. Στο παρακάτω τμήμα από ένα αρχείο ped, ποιος είναι ο πατέρας του ατόμου GR00103;

GR001	GR00101	0	0	1	1	.....
GR001	GR00102	0	0	2	1	.....
GR001	GR00103	GR00101	GR00102	2	2	.....
GR002	GR00201	0	0	1	1	.....
GR002	GR00202	0	0	2	1	.....
GR002	GR00203	GR00201	GR00202	1	2	.....



29. Το άτομο GR00202 είναι γυναίκα ή άνδρας; Είναι ασθενής ή υγιής;

30. Στο παράθυρο του HAPLOVIEW, πατήστε OK.

31. Τι ονομάζεται **ετεροζυγωτία** στη Γενετική Πληθυσμών;

32. Ανοίγει η καρτέλα Check Markers (Εικόνα 6). Εδώ παρουσιάζονται γενικά στατιστικά στοιχεία για τους δείκτες που χρησιμοποιούνται στην ανάλυση. Εμφανίζεται η **παρατηρούμενη (ObsHET)** και η **αναμενόμενη (PredHET)** ετεροζυγωτία σύμφωνα με την ισορροπία Hardy-Weinberg (ποσοστό ετεροζυγωτών στον πληθυσμό) καθώς και ο στατιστικός έλεγχος της απόκλισης (HWpval). Ποιες αναλογίες γονοτύπων προβλέπει η ισορροπία Hardy-Weinberg σε έναν πληθυσμό;

#	Name	Position	ObsHET	PredHET	HWpval	%Geno	FamTrio	MendErr	MAF	Alleles	Rating
1	C_506524_10	0	0.453	0.46	1.0	98.1	0	0	0.358	C:A	✓
2	C_442248_10	3951	0.463	0.461	1.0	100.0	0	0	0.361	C:A	✓
3	C_453657_10	8664	0.453	0.46	1.0	98.1	0	0	0.358	C:A	✓
4	C_293600_10	13993	0.259	0.278	0.8805	100.0	0	0	0.167	A:C	✓
5	C_11600662_10	25599	0.463	0.461	1.0	100.0	0	0	0.361	C:A	✓
6	C_25958629_10	31689	0.208	0.269	0.218	98.1	0	0	0.16	A:C	✓
7	C_27949589_10	32947	0.259	0.252	1.0	100.0	0	0	0.148	A:C	✓
8	C_11600261_10	39682	0.315	0.356	0.5722	100.0	0	0	0.231	A:C	✓
9	C_308042_10	67220	0.283	0.34	0.366	98.1	0	0	0.217	A:C	✓
10	C_40609_10	122540	0.635	0.5	0.1076	95.3	0	0	0.49	C:A	✓
11	C_11600340_10	130959	0.667	0.499	0.032	100.0	0	0	0.481	C:A	✓
12	C_315670_10	135784	0.415	0.406	1.0	98.1	0	0	0.283	C:A	✓
13	C_26959079_10	162236	0.426	0.424	1.0	100.0	0	0	0.306	C:A	✓
14	C_12042882_20	192006	0.352	0.356	1.0	100.0	0	0	0.231	A:C	✓
15	C_7968518_10	291351	0.389	0.45	0.4358	100.0	0	0	0.343	C:A	✓
16	C_7968314_10	303610	0.5	0.424	0.3557	100.0	0	0	0.306	A:C	✓
17	C_27491053_10	328167	0.352	0.375	0.8478	100.0	0	0	0.25	C:A	✓
18	C_7969752_1_	367814	0.481	0.497	0.9771	100.0	0	0	0.463	A:C	✓
19	C_11258596_10	373318	0.481	0.497	0.9771	100.0	0	0	0.463	A:C	✓
20	C_7460329_20	386530	0.426	0.393	0.8579	100.0	0	0	0.269	A:C	✓
21	C_1180387_10	400653	0.481	0.444	0.8205	100.0	0	0	0.333	A:C	✓
22	C_1180376_20	412735	0.396	0.397	1.0	98.1	0	0	0.274	A:C	✓
23	C_11258491_20	421906	0.5	0.471	0.9291	100.0	0	0	0.38	A:C	✓
24	C_1180364_1_	430250	0.574	0.496	0.4113	100.0	0	0	0.454	A:C	✓
25	C_1180346_10	443725	0.389	0.313	0.1774	100.0	0	0	0.194	A:C	✓
26	C_1180342_10	449664	0.222	0.431	0.001	100.0	0	0	0.315	A:C	✓
27	C_1180338_20	455092	0.396	0.38	1.0	98.1	0	0	0.255	C:A	✓
28	C_15914229_10	475666	0.5	0.486	1.0	100.0	0	0	0.417	A:C	✓

Εικόνα 6. Καρτέλα Check Markers

.....  
.....  
33. Στα δεδομένα που αναλύονται, υπάρχει κάποιο SNP που δεν έχει γονοτυπικές αναλογίες στον πληθυσμό σύμφωνα με την ισορροπία Hardy-Weinberg; Ποιο είναι αυτό; Παρατηρούνται λιγότεροι ή περισσότεροι ετεροζυγώτες από το αναμενόμενο;

.....  
.....  
34. Τι σημαίνει όταν ένας γενετικός δείκτης δεν ικανοποιεί την ισορροπία Hardy-Weinberg;

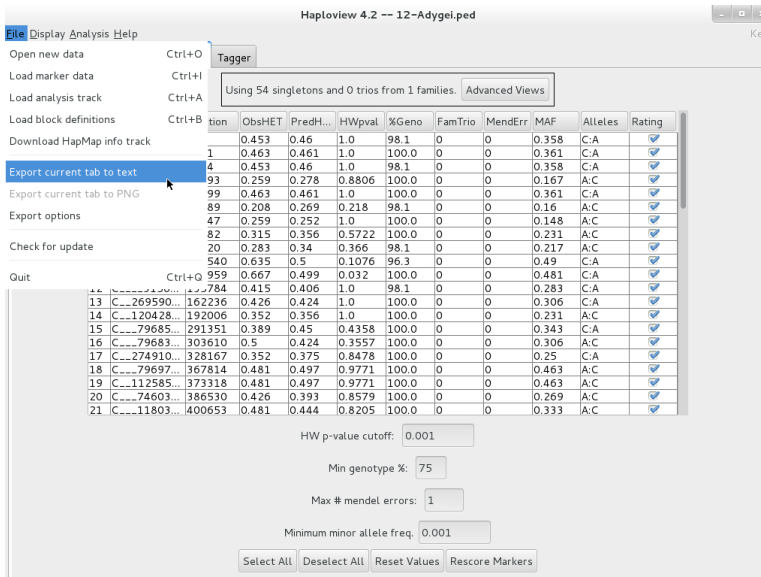
.....  
.....  
35. Στη στήλη “%Geno” φαίνεται το ποσοστό επιτυχίας της δοκιμασίας γονοτύπησης για το κάθε SNP (για πόσα άτομα “πέτυχε” η δοκιμασία).

36. Στην προτελευταία στήλη φαίνεται η συχνότητα του πιο σπάνιου αλληλομόρφου για κάθε SNP (minor allele frequency – MAF).

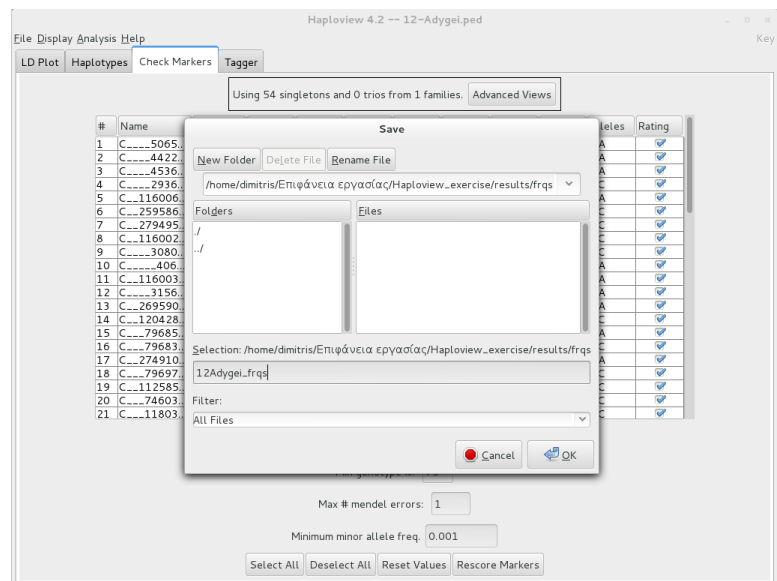
37. Ποιο SNP είναι το λιγότερο πολυμορφικό;

.....  
.....  
38. Στην τελευταία στήλη μπορείτε να αποφασίσετε αν θέλετε, μετά από αυτόν τον ποιοτικό έλεγχο, να απορρίψετε κάποια SNPs από άλλες αναλύσεις. Εδώ θα αναλύσουμε όλα τα SNPs. Βεβαιωθείτε ότι είναι επιλεγμένα όλα τα SNPs.

39. Σώστε τα αποτελέσματα, επιλέγοντας **File → Export current tab to text**, στον υποφάκελο **frqs** του φακέλου **results** με **όνομα αρχείου 12adygei\_frqs** (Εικόνες 7α, 7β) (στη συνέχεια θα επαναλάβετε την ίδια διαδικασία για όλους τους πληθυσμούς. Δώστε προσοχή στη **σωστή ονομασία των αρχείων** που σώζετε. **ΟΛΑ ΤΑ ΓΡΑΜΜΑΤΑ ΜΙΚΡΑ, χωρίς κενά, προσοχή στην επέκταση και μπροστά από το όνομα του πληθυσμού υπάρχει ένα διψήφιο νούμερο, όπως φαίνεται στο παράρτημα στο τέλος των οδηγιών**).

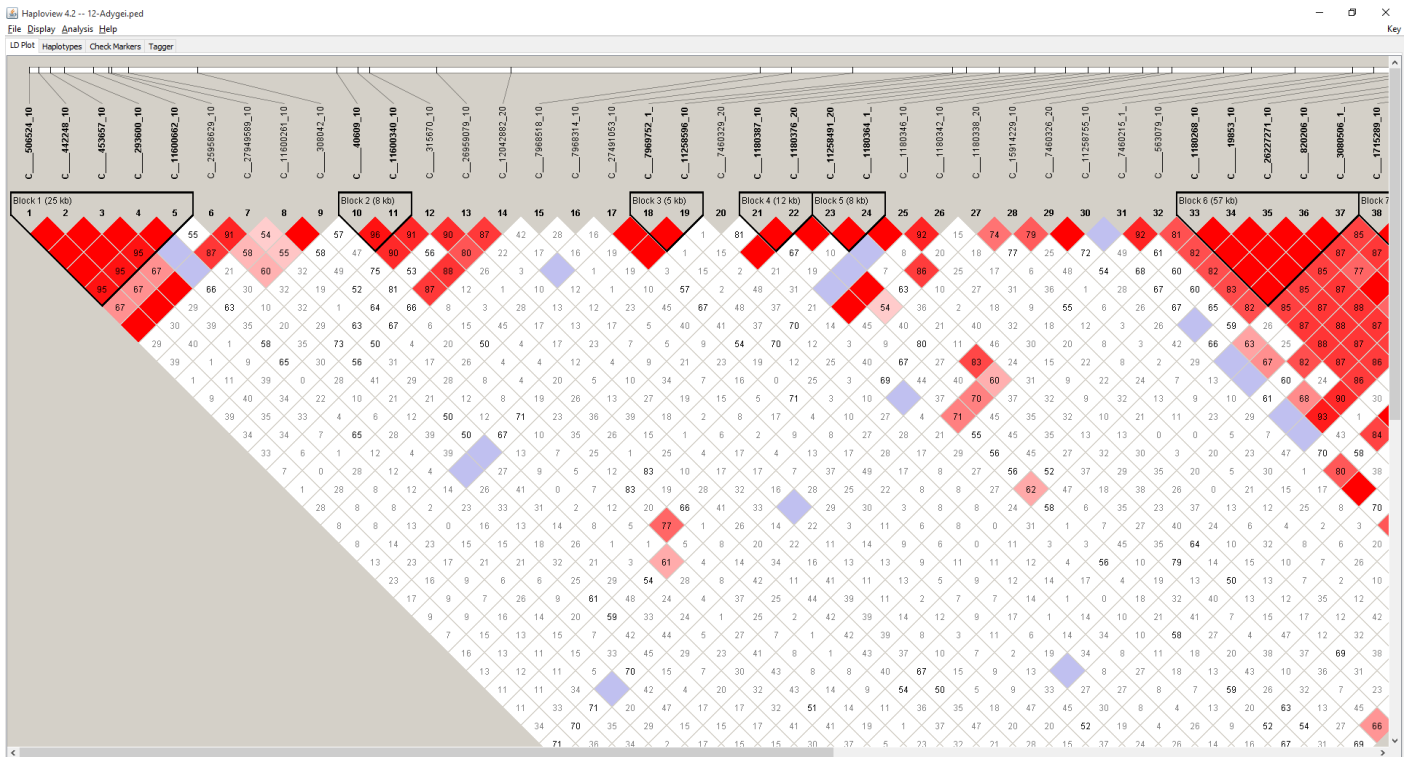


Εικόνα 7α. File → Export current tab to text



Εικόνα 7β. Αποθήκευση στο φάκελο frqs

40. Επιλέξτε την καρτέλα **LD plot**. Παρατηρήστε το πρότυπο ανισορροπίας σύνδεσης για την περιοχή και τον συγκεκριμένο πληθυσμό (Εικόνα 8).



Εικόνα 8. Καρτέλα LD Plot

41. Κάθε τετράγωνο αντιπροσωπεύει την τιμή της παραμέτρου ανισορροπίας σύνδεσης (σε αυτή την περίπτωση του D') ανάμεσα σε ένα ζεύγος SNPs. Τι τιμές μπορεί να πάρει το D';

.....

42. Ποιο είναι το D' ανάμεσα στα SNPs C\_11600261\_10 και C\_40609\_10;

.....

43. Το χρώμα αντιστοιχεί σε “ένταση” του D' καθώς και σε **στατιστική σημαντικότητα του αποτελέσματος** (μετρημένη με ένα LOD score). Το D' είναι πολύ ευαίσθητο στο μέγεθος του δείγματος, οπότε σε μικρά δείγματα, μπορεί να προκύψει υψηλό D' που να μην είναι στατιστικά σημαντικό (στην περίπτωση ελέγχου ενός μεγαλύτερου δείγματος, η συσχέτιση θα εξαφανιστεί). Αυτό ελέγχεται με το LOD score. Ψάχνουμε **LOD score μεγαλύτερο ή ίσο με 2** για να μπορούμε να θεωρήσουμε το αποτέλεσμά μας αξιόπιστο (Εικόνα 9). Αν το **LOD δεν είναι υψηλό**, τότε το τετράγωνο παραμένει λευκό (Εικόνα 10). Με μπλε σημειώνονται τα ζευγάρια των SNPs για τα οποία το D' είναι ίσο με 1 αλλά το LOD score είναι κάτω από 2 (Εικόνα 11).

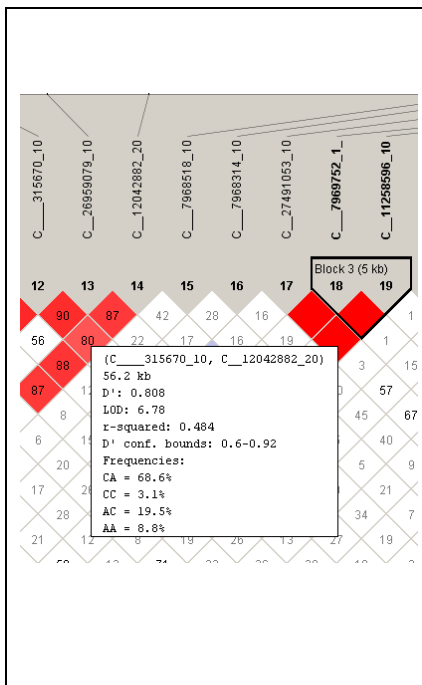
44. Η εμφάνιση λεπτομερών στοιχείων αναφορικά με την ανισορροπία σύνδεσης μεταξύ δυο SNPs μπορεί να γίνει με παρατεταμένο δεξί κλικ επάνω σε ένα τετράγωνο (Εικόνα 9, Εικόνα 10, Εικόνα 11).

### **LOD score**

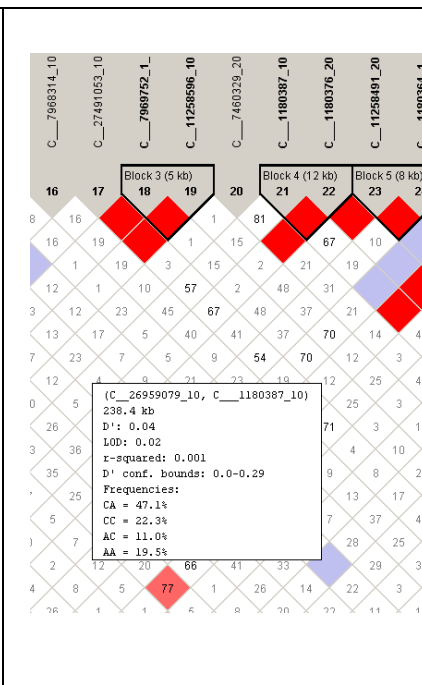
Το **LOD score** προέρχεται από τις λέξεις *Logarithm Of Odds*. Ο λογάριθμος έχει βάση το 10.

Το **LOD score** παρέχει μια στατιστική εκτίμηση του κατά πόσο δυο τόποι (*loci* – οι θέσεις των γονιδίων) είναι πιθανό να βρίσκονται κοντά ο ένας με τον άλλο σε ένα χρωμόσωμα και συνεπώς, να συν κληρονομούνται σαν ένα πακέτο.

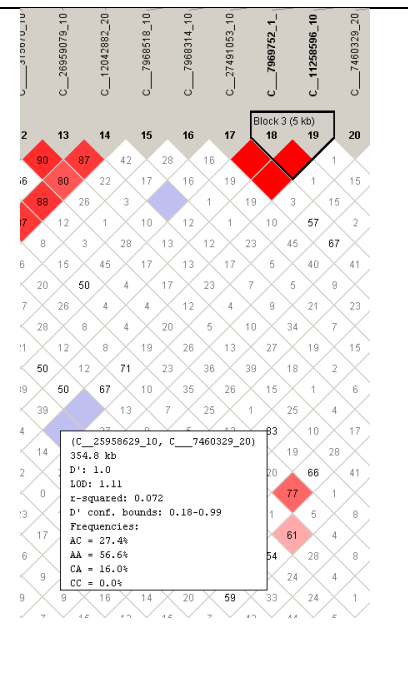
Ένα **LOD score** με τιμή 3, σημαίνει ότι οι πιθανότητες είναι μια στις χίλιες υπέρ της γενετικής σύνδεσης.



**Εικόνα 9.** Λεπτομέρειες ανισορροπίας σύνδεσης δυο SNPs με  $LOD \geq 2$



**Εικόνα 10.** Λεπτομέρειες ανισορροπίας σύνδεσης δυο SNPs με  $LOD < 2$

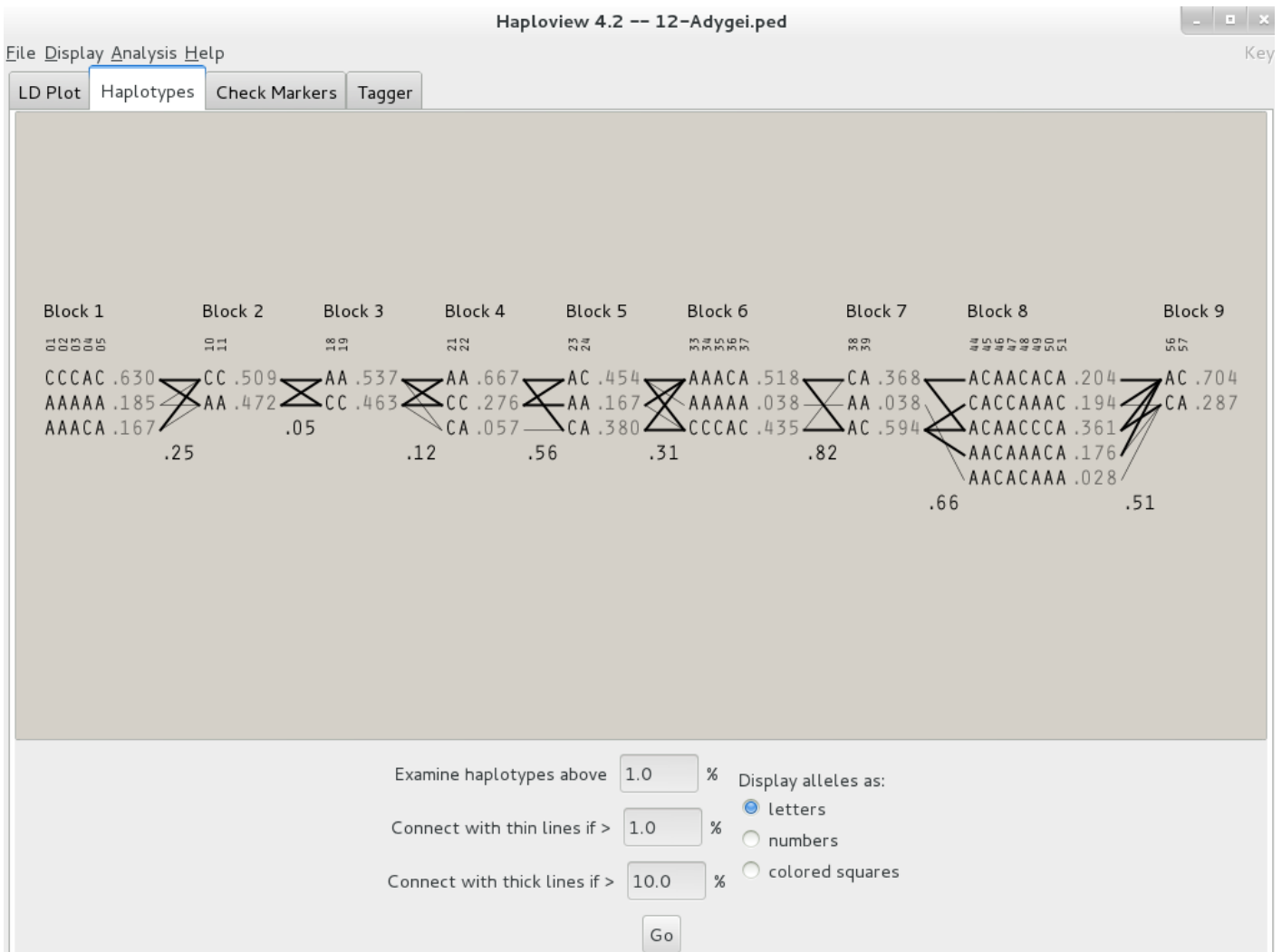


**Εικόνα 11.** Λεπτομέρειες ανισορροπίας σύνδεσης δυο SNPs με  $LOD < 2$  και  $D' = 1$

45. Το HAPLOVIEW επιλέγει τις ομάδες των SNPs για τις οποίες υπάρχει πολύ υψηλή συσχέτιση, ώστε να δημιουργηθούν “blocks”. Τα blocks αυτά δεν έχουν πολλούς συνδυασμούς αλληλομόρφων (απλότυπους) και μπορεί να θεωρηθεί ότι συμβαίνει σπάνια ανασυνδυασμός ανάμεσά τους. Ποιο είναι το μεγαλύτερο block που παρατηρείτε εδώ (δώστε τα νούμερα των SNPs).

46. Αν τα SNPs αυτά δεν ήταν σε ανισορροπία σύνδεσης (δεν συσχετίζονταν δηλαδή), πόσους απλότυπους θα μπορούσαμε να παρατηρήσουμε στο δείγμα μας;

47. Στην καρτέλα **Haplotypes** παρουσιάζονται οι απλότυποι που “πραγματικά” παρατηρούνται στο δείγμα μας (εντός εισαγωγικών γιατί οι απλότυποι αυτοί προβλέπονται στατιστικά και όχι πειραματικά) (Εικόνα 12). Πόσοι απλότυποι παρατηρούνται σε αυτόν τον πληθυσμό για τη συγκεκριμένη ομάδα SNPs;



**Εικόνα 12.** Καρτέλα Haplotypes

48. Σημείωση: Τα αλληλόμορφα στους απλότυπους φαίνονται όλα σαν AC, γιατί στα δεδομένα που εισάγαμε στο πρόγραμμα, οι γενότυποι είχαν κωδικοποιηθεί με νούμερα (ως 12 - 11 - 22). Επιλέξτε **“Display alleles as: numbers”** και **“Go”** για να δείτε τα δεδομένα όπως πραγματικά τα δώσαμε.

49. Δίπλα σε κάθε απλότυπο, εμφανίζεται η συχνότητά του στον πληθυσμό. Για τη συγκεκριμένη ομάδα SNPs, ποιος απλότυπος είναι ο πιο συχνός και με τι συχνότητα;

.....

50. Επιστρέψτε στην καρτέλα του LDplot. Από το μενού Display → LD zoom μπορείτε να κάνετε zoom in και zoom out. Μπορείτε επίσης να δείτε το αντίστοιχο σχήμα για τις τιμές του  $r^2$ , **Display → LD color scheme → R squared**. Το  $r^2$  είναι πολύ πιο αυστηρό από το D' και όπως βλέπετε, πολύ λιγότερα ζευγάρια εμφανίζεται να συσχετίζονται.

51. Σώστε τα αποτελέσματα στον υποφάκελο **lds** του φακέλου **results**, ακολουθώντας *File* → *Export current tab to text* πχ. *12adygei\_ld*, όπως στο *Βήμα 43*. **ΟΠΩΣ ΚΑΙ ΠΡΙΝ ΔΩΣΤΕ ΠΡΟΣΟΧΗ ΣΤΟ ΟΝΟΜΑ ΤΟΥ ΑΡΧΕΙΟΥ**. Μπροστά από το όνομα του πληθυσμού υπάρχει ένα διψήφιο **νούμερο**, όπως φαίνεται στο παράρτημα στο τέλος των οδηγιών).
52. **Σημείωση**: μπορείτε να σώσετε είτε την εικόνα (File → Export current tab to PNG) ή τον πίνακα με όλα τα αποτελέσματα της ανάλυσης.
53. Πηγαίνετε στην καρτέλα **Tagger**. Από εδώ μπορείτε να επιλέξετε tagging SNPs (δηλαδή SNPs που αντιπροσωπεύουν μια ολόκληρη ομάδα ενός block. Οι γενότυποι αυτών των SNPs μπορούν να χρησιμοποιηθούν για να προβλέψουμε τους γενότυπους και στα άλλα SNPs της ομάδας. Με άλλα λόγια, είναι περιττό να τυποποιήσουμε όλα τα SNPs μιας ομάδας. Παίρνουμε την ίδια πληροφορία τυποποιώντας μόνο τα tagging SNPs.
54. Ο αλγόριθμος που χρησιμοποιείται για την επιλογή tagging SNPs από το HAPLOVIEW, βασίζεται στον υπολογισμό του  $r^2$ . Συνοπτικά, υπολογίζονται όλα τα κατά ζεύγη τεστ και επιλέγονται ως αντιπροσωπευτικά, τα SNPs που “πιάνουν” ένα άλλο με  $r^2$  ίσο ή μεγαλύτερο από κάποιο όριο ( $r^2$  threshold). Συνήθως, το όριο αυτό είναι 0.8, μπορείτε να το αλλάξετε από το αντίστοιχο πεδίο.
55. Επιλέξτε **pairwise tagging only**. Με τις άλλες επιλογές δίνεται η δυνατότητα να επιλεγούν απλότυποι 2 ή 3 SNPs ως αντιπροσωπευτικοί και δίνονται καλύτερα αποτελέσματα (μεγαλύτερη οικονομία). Επιλέξτε **“Run Tagger”**.
56. Πηγαίνετε στην καρτέλα **results**. Από την αριστερή μεριά, στον πάνω πίνακα εμφανίζονται όλα τα SNPs που μελετάμε. Επιλέξτε για παράδειγμα το SNP C\_820206\_10. Στον κάτω πίνακα φαίνονται τα SNPs που αυτό το SNP μπορεί να αντιπροσωπεύσει (tagged SNPs), συμπεριλαμβανομένου και του ίδιου. Πόσα SNPs “πιάνει” το C\_820206\_10;
- .....
57. Επιλέξτε το SNP C\_27491053\_10. Τι ισχύει για αυτό το SNP;
- .....

58. Στα αριστερά, κάτω από τους πίνακες, αναγράφεται ο αριθμός των SNPs που αναλύθηκαν καθώς και ο αριθμός των tagging SNPs που απαιτούνται. Πόσα tagging SNPs χρειάζονται από τα 63 που μελετήθηκαν;

59. Στον πίνακα στα δεξιά, φαίνονται όλα τα τεστ ανισορροπίας σύνδεσης. Το SNP C\_442248\_10 ποια άλλα SNPs μπορεί να αντιπροσωπεύσει και με τι  $r^2$ ;

60. Επαναλάβετε τη διαδικασία ανάλυσης από την αρχή για **δύο πληθυσμούς από κάθε γεωγραφική περιοχή** (κάθε φορά σώζοντας τα αποτελέσματα από την καρτέλα CHECK MARKERS με όνομα αρχείου **#πληθυσμός\_frqs** στον φάκελο **frqs** και τα αποτελέσματα από την καρτέλα LDplot με **#πληθυσμός\_ld** στον φάκελο **lds**). Κάθε φορά ελέγξτε αν ισχύει η ισορροπία Hardy-Weinberg για όλους τους δείκτες, **αλλά φροντίστε να τους συμπεριλάβετε όλους στην ανάλυση που κάνετε εδώ**. Χρησιμοποιήστε τα ονόματα των πληθυσμών με τα νούμερα που φαίνονται στο παράρτημα. **Πάντα** σώζετε στους υποφακέλους του φακέλου **results**, σύμφωνα με τις οδηγίες του βήματος αυτού.

61. Πηγαίνετε στους φακέλους **frqs** και **lds** και δείτε τη λίστα με τα αρχεία που σώσατε.

62. Ανοίξτε τα αρχεία “**frqs**” και “**ld**” που μόλις δημιουργήσατε χρησιμοποιώντας την εφαρμογή LibreOffice Calc. Αυτό επιτυγχάνεται με δεξί κλικ σε κάθε αρχείο. Αν δεν εμφανίζεται η συγκεκριμένη εφαρμογή στις διαθέσιμες δυνατότητες, τότε επιλέγουμε **Open with Other Application...** και στο νέο παράθυρο βρίσκουμε την εφαρμογή LibreOffice Calc. Για τη σωστή εμφάνιση των αποτελεσμάτων επιλέγουμε τις δυνατότητες: Merge delimiters και Space. Παρατηρήστε τη μορφή των αρχείων αυτών.

Τα αρχεία αυτά βρίσκονται στους φακέλους **Haploview\_exercise/results/frqs** και **Haploview\_exercise/results/lds**

**ΠΑΡΑΤΗΡΗΣΗ 1:** Στην περίπτωση που δεν υπάρχει το LibreOffice Calc, μπορείτε με διπλό κλικ να ανοίξετε τα αρχεία με τον κειμενογράφο

**ΠΑΡΑΤΗΡΗΣΗ 2:** Στην περίπτωση λειτουργικού συστήματος Windows και της εφαρμογής Microsoft Office, θα μπορούσατε να δείτε τα αρχεία με κατάληξη **ped**, χρησιμοποιώντας το Excel.



63. Φτιάξτε γραφική παράσταση της συχνότητας των σπανίων αλληλομόρφων (MAF-minor allele frequency) για τα SNPs 1, 10, 15, 20, 36, 52, 57 σε όλους τους πληθυσμούς (πχ. για το SNP C\_506524\_10). Τις τιμές MAF θα τις αναζητήσετε στα αρχεία \_frqs που αποθηκεύσατε.

64. Δημιουργήστε ένα καινούργιο αρχείο excel στην επιφάνεια εργασίας με τίτλο “MAF\_Figure”, στο οποίο θα συγκεντρώσετε τις τιμές MAF των SNPs σε έναν πίνακα με τη μορφή:

Πληθυσμοί	Πληθ.Α	Πληθ.Β	...
SNP1	MAF	MAF	...

65. Για να δημιουργήσετε το διάγραμμα επιλέξτε ολόκληρο τον πίνακα. Στη συνέχεια στην καρτέλα Insert, διαλέξτε Column και την πρώτη επιλογή στα 2-D Columns. Επαναλάβετε τη διαδικασία και για τα υπόλοιπα SNPs.

66. Περιγράψτε τι παρατηρείτε για το κάθε SNP και τους διαφορετικούς πληθυσμούς.

.....

.....

.....

.....

.....

.....

67. Σε ποιους παράγοντες μπορεί να οφείλονται τα αποτελέσματα που παρατηρείτε;

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

68. Φτιάξτε γραφική παράσταση της τιμής  $r^2$  για τα ζεύγη των SNPs 1-2, 18-19, 40-41, 62-63 σε όλους τους πληθυσμούς. Δημιουργήστε ένα καινούργιο αρχείο excel στην επιφάνεια εργασίας σας με τίτλο “r2\_figure”. Ο πίνακας για το διάγραμμα θα πρέπει να είναι της μορφής που ακολουθεί:

SNP1-SNP2	$r^2$ values
ΠληθυσμόςΑ	
ΠληθυσμόςΒ	
...	

69. Για να δημιουργήσετε το διάγραμμα επιλέξτε ολόκληρο τον πίνακα. Στη συνέχεια στην καρτέλα Insert, διαλέξτε Column και την πρώτη επιλογή στα 2-D Columns. Επιλέξτε τον πίνακα και έπειτα δεξί κλικ και Select Data. Στο παράθυρο που ανοίγει επιλέξτε το “r<sup>2</sup>” και πατήστε Edit (ακριβώς από πάνω). Στο “Series name” βάλτε το όνομα του πληθυσμού και στο “Series values” σβήστε τα δεδομένα που υπάρχουν και επιλέξτε την περιοχή που εμφανίζονται οι τιμές  $r^2$  για τον πρώτο πληθυσμό της επιλογής σας. Πατήστε OK.

70. Αφού παρατηρήσετε το διάγραμμα, επαναλάβετε την ίδια διαδικασία και για τα υπόλοιπα ζεύγη SNPs.

71. Περιγράψτε τι παρατηρείτε και συγκρίνετε τα αποτελέσματα μεταξύ τους.

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

.....

*Για τη γραπτή αναφορά σας, επαναλάβετε τα βήματα 63-67 για δύο (2) από τα SNPs που αναγράφονται στο βήμα 63, καθώς και τα βήματα 68-71 για ένα (1) από τα ζεύγη των SNPs που αναφέρονται στο βήμα 68.*

## Παράρτημα

	<b>SNPs</b>	<b>POPULATIONS STUDIED</b>
1	C___506524_10	<u>AFRICA</u>
2	C___442248_10	01Biaka
3	C___453657_10	02Mbuti
4	C___293600_10	03Yoruba
5	C__11600662_10	04Ibo
6	C__25958629_10	05Hausa
7	C__27949589_10	06Chagga
8	C__11600261_10	07Ethiopians
9	C___308042_10	08African_Americans
10	C___40609_10	
11	C__11600340_10	<u>SW ASIA</u>
12	C___315670_10	09Yemenites
13	C__26959079_10	10Druze
14	C__12042882_20	11Samaritans
15	C___7968518_10	
16	C___7968314_10	<u>EUROPE</u>
17	C__27491053_10	12Adygei
18	C___7969752_1_	13Chuvash
19	C__11258596_10	14Ashkenazi
20	C___7460329_20	15Finns
21	C___1180387_10	16Danes
22	C___1180376_20	17Irish
23	C__11258491_20	18European_Americans
24	C___1180364_1_	
25	C___1180346_10	<u>EAST ASIA</u>
26	C___1180342_10	19Khanty
27	C___1180338_20	20KomiZyriane
28	C__15914229_10	21San Francisco Chinese
29	C___7460326_20	22Taiwanese_Chinese
30	C__11258755_10	23Hakka
31	C___7460215_1_	24Japanese
32	C___563079_10	25Ami
33	C___1180268_10	26Atayal
34	C___19853_10	27Cambodians
35	C__26227271_10	
36	C___820206_10	<u>PACIFIC-SIBERIA</u>
37	C___3080506_1_	28Yakut
38	C___1715289_10	29Nasioi
39	C___820226_10	30Micronesians
40	C__11258355_10	
41	C___201617_10	<u>NORTH AMERICA-SOUTH AMERICA</u>
42	C__26227314_10	31Cheyenne

43	C__1674439_10	32Pima_Arizona
44	C__1674429_10	33Pima_Mexico
45	C__1674422_1_	34Maya
46	C__1674419_10	35Ticuna
47	C__1674406_10	36Karitiana
48	C__1674398_10	37Rondonian_Surui
49	C_15961408_10	
50	C__1674359_10	
51	C__7460104_1_	
52	C_15969780_10	
53	C__1674327_10	
54	C__1674292_30	
55	C__1674268_10	
56	C__1674257_10	
57	C_11258496_10	
58	C_28036504_10	
59	C____33113_10	
60	C_26006993_10	
61	C_26068473_10	
62	C_11225040_10	
63	C____199633_10	

—