

Hapl-o-Mat - Getting Started Windows

Please also see the README.

Hapl-o-Mat

Hapl-o-Mat is software for HLA haplotype inference coded in C++. Besides estimating haplotype frequencies via an expectation-maximization algorithm, it is capable of processing HLA genotype population data. This includes translation of alleles between various typing resolutions and resolving allelic and genotypic ambiguities. Both common formats for recording HLA genotypes, multiple allele codes (MAC) and genotype list strings (GLS), are supported.

This guide explains how to use Hapl-o-Mat under Windows. For more information refer to our publications on Hapl-o-Mat:

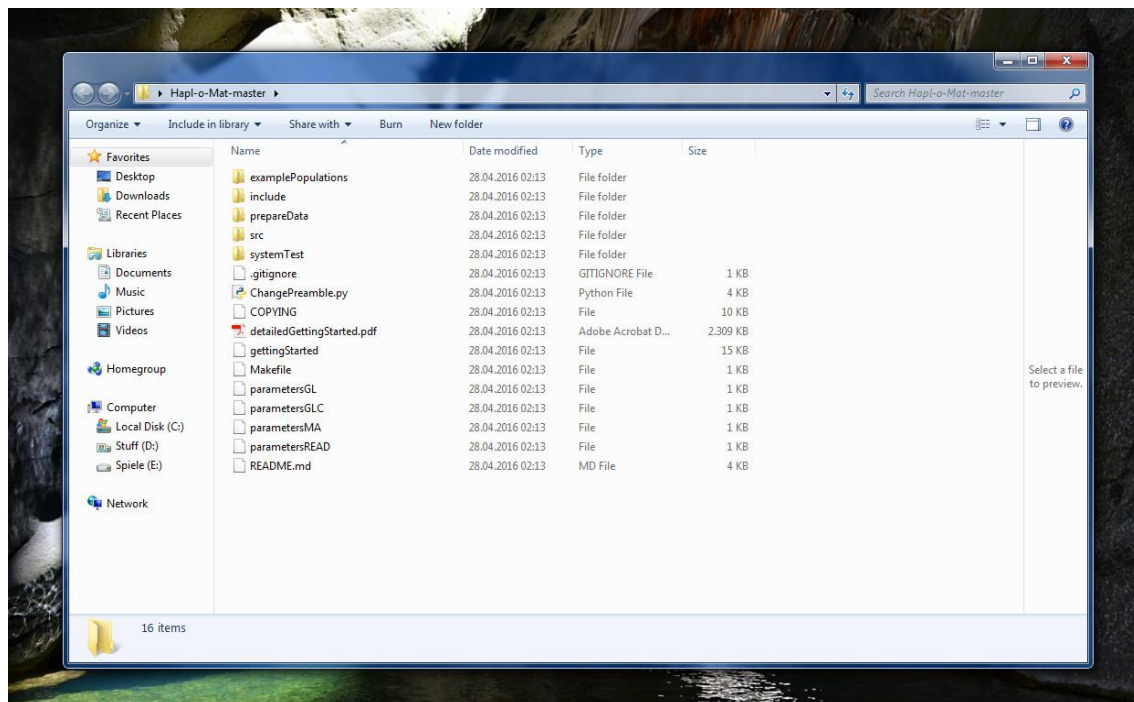
Journal article to come

C. Schaefer, A.H. Schmidt, J. Sauter: Hapl-O-mat: A Versatile Software for Haplotype Frequency Estimation. HLA (2016), 87, 236-320

Getting Started

This guide is an introduction on how to use Hapl-o-Mat. In order to follow this guide, you need a Windows system and a C++ compiler supporting C++11. In this tutorial, we use Eclipse IDE for C/C++ Developers. You can get it here <https://www.eclipse.org/downloads/>.

After successfully downloading Hapl-o-Mat, look into the folder. You should see the following:



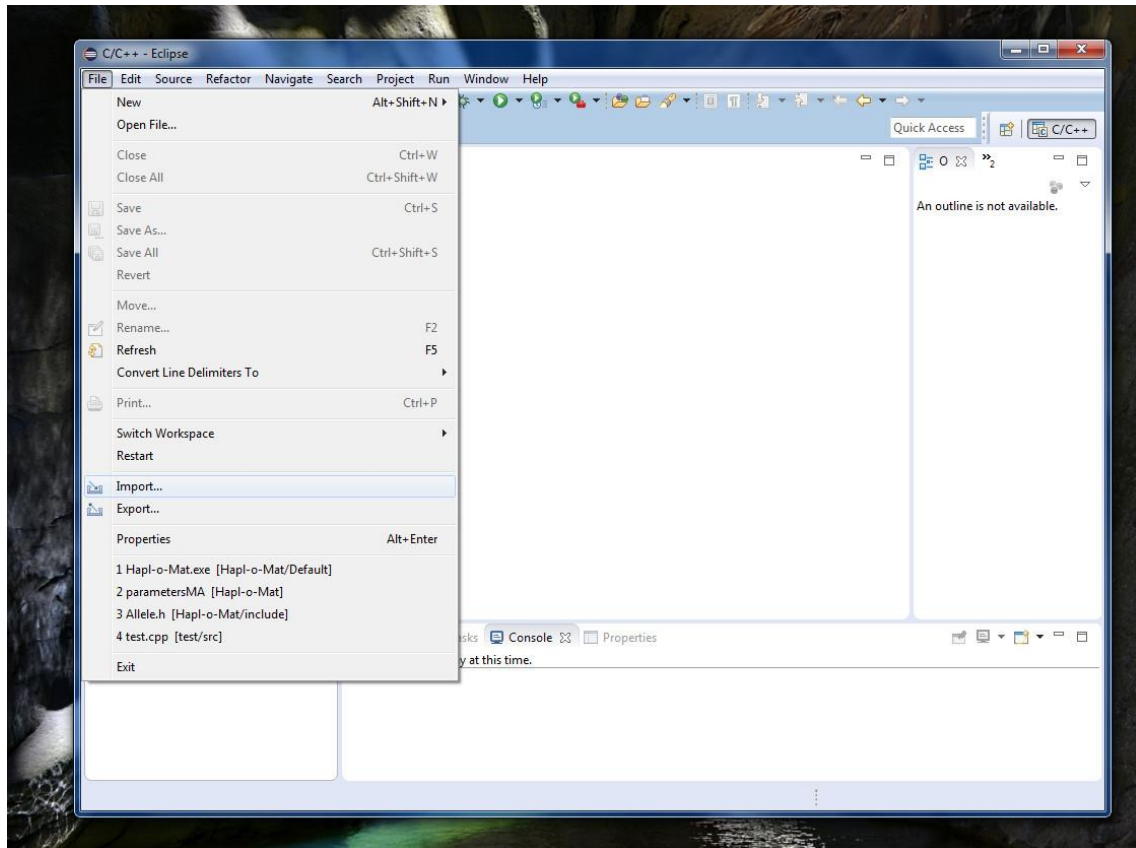
We give you some information on the files you find. Files which are important for using Hapl-o-Mat are marked as bold:

File name	Description
ChangePreamble.py	A python script to adapt the preamble in all files. You are not going to use it.
COPYING	The GNU General Public License
detailedGettingStartedLinux	Guide for using Hapl-o-Mat under Linux
detailedGettingStartedWindows	Guide for using Hapl-o-Mat under Windows
examplePopulations	Some genotype population data we are going to work with in the section Tutorials.
gettingStarted	A shorter form of the Linux tutorial
include	A part of Hapl-o-Mat's source code; If you do not want to change code, do not touch it.
Makefile	Instructions for building Hapl-o-Mat
parametersGLS, parametersGLSC, parametersMAC, parametersREAD	Parameter files for Hapl-O-mat; We are going to discuss this in section Parameters.
prepareData	Here is everything to create the data required by Hapl-o-Mat.
README	Read me
src	A part of Hapl-o-Mat's source code; If you do not want to change code, do not touch it.
systemTest	Run the system test after changing code to check, if you broke something. Refer to its README.

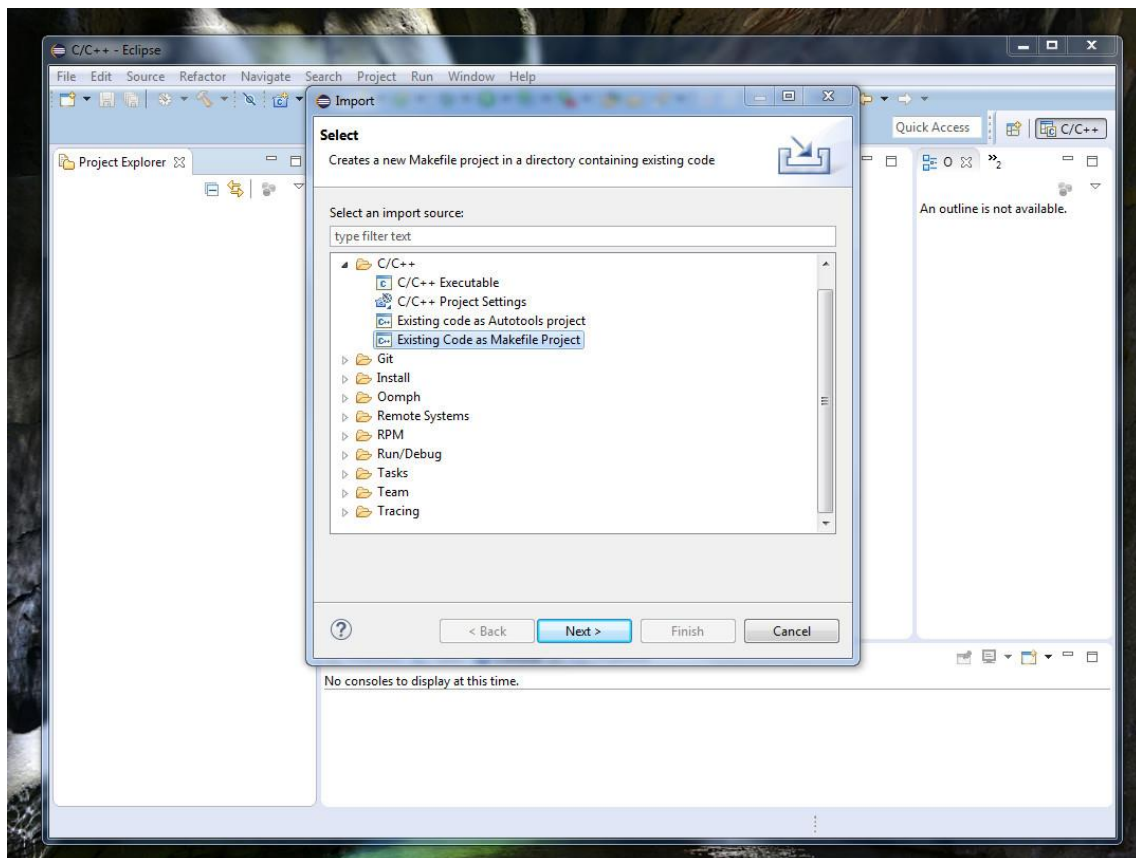
To estimate haplotype frequencies we only need to consider the folder prepareData and the files parametersGLS, parametersGLSC, parametersMAC, and parametersREAD. To finish this tutorial we need the folder examplePopulations, too.

Install Hapl-o-Mat

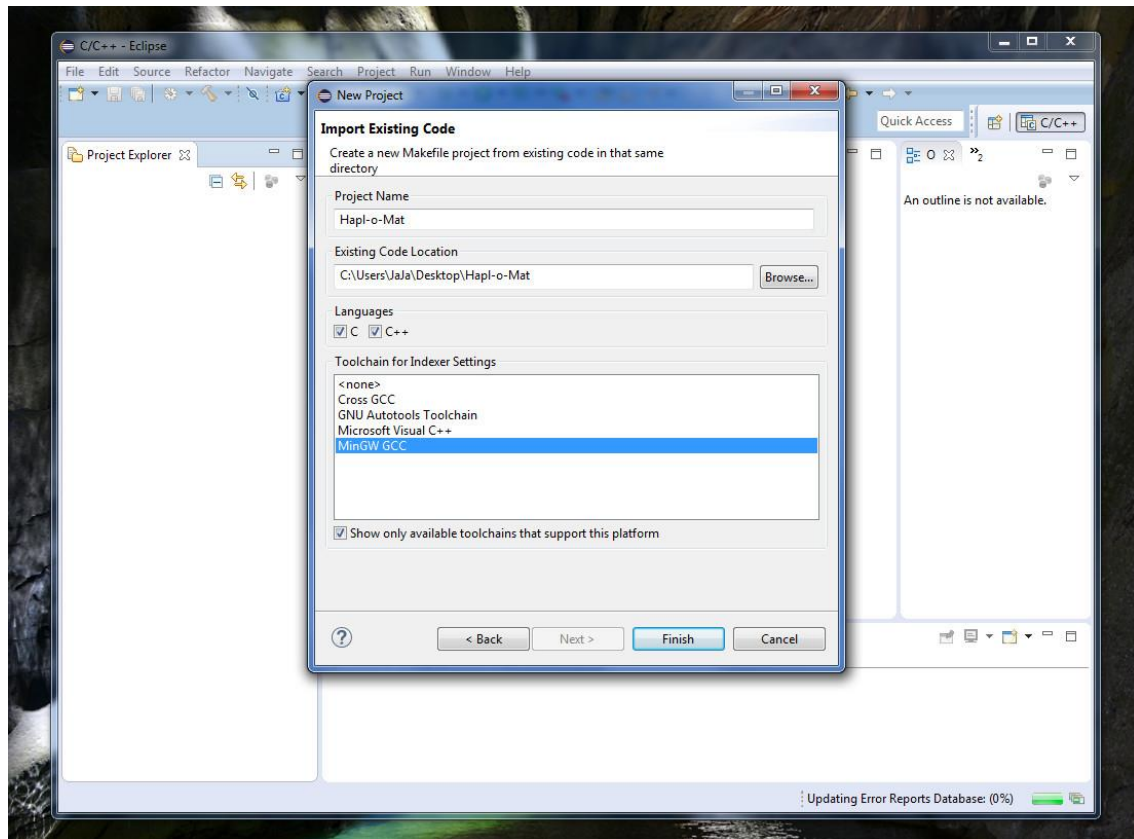
We compile Hapl-o-Mat using Eclipse. Start Eclipse and create a project by clicking on File -> Import.



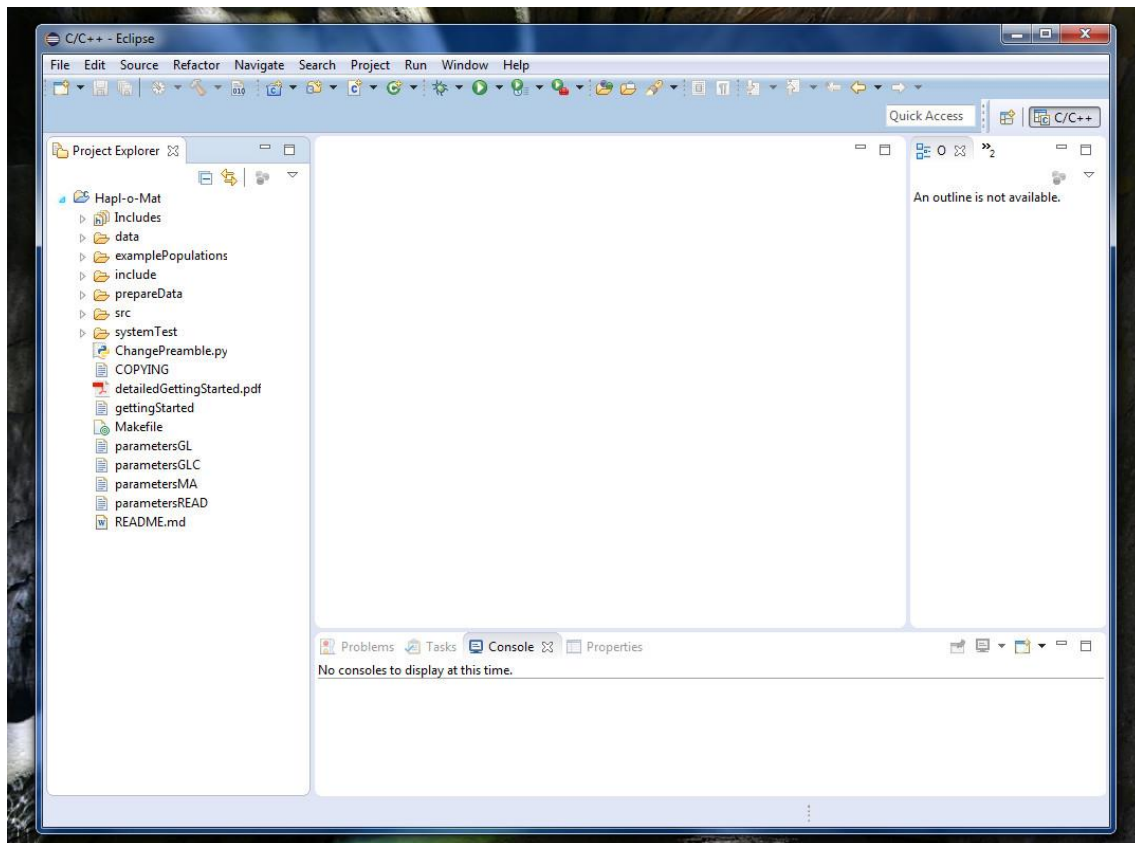
Choose Existing Code as Makefile Project in the pop-up window and then click Next.



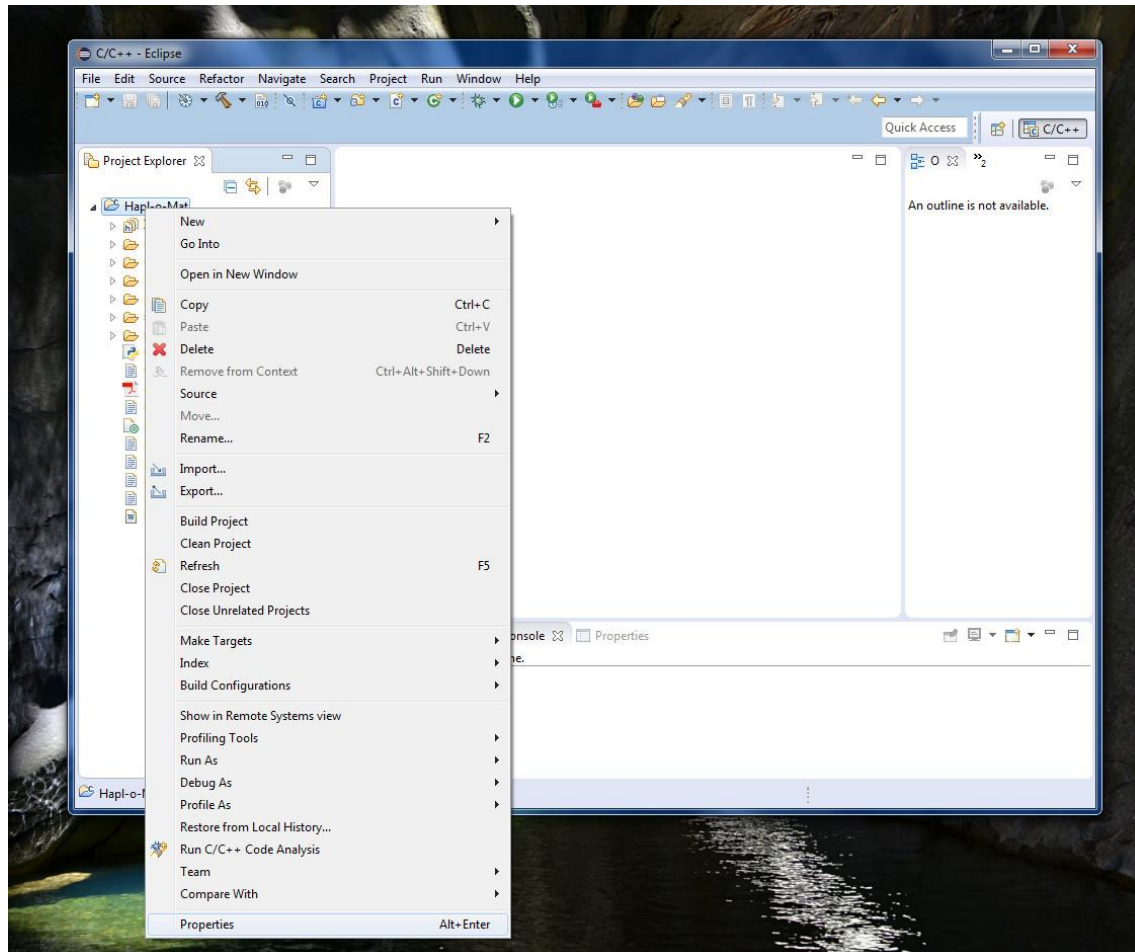
Enter the Project Name, e.g. Hapl-o-Mat, and browse to the location where you saved Hapl-o-Mat. Select MinGW GCC as Toolchain. Click Finish.



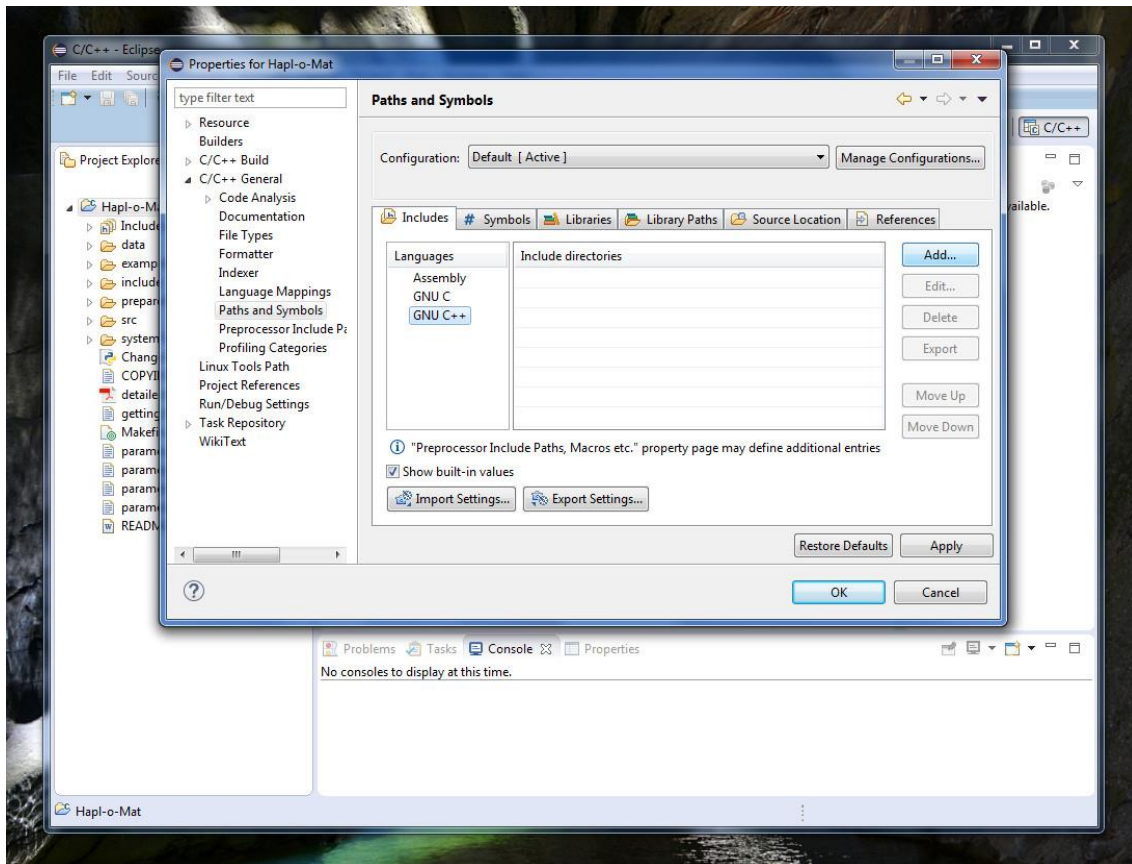
Now, you find all files and folders of Hapl-o-Mat in the Project Explorer. If your Eclipse does not show the Project Explorer, activate it via Window -> Show View -> Project Explorer.



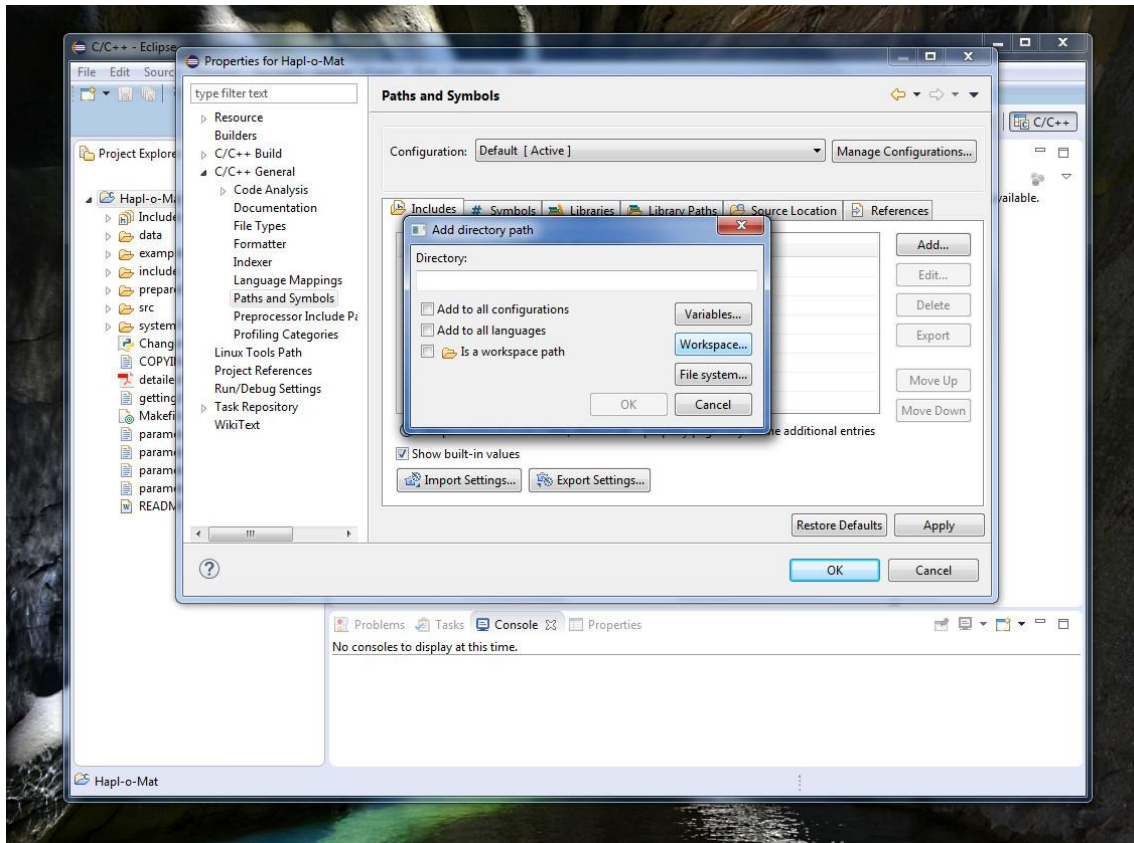
Next, we have to adapt some settings. Right-click into the Project Explorer and choose Properties.



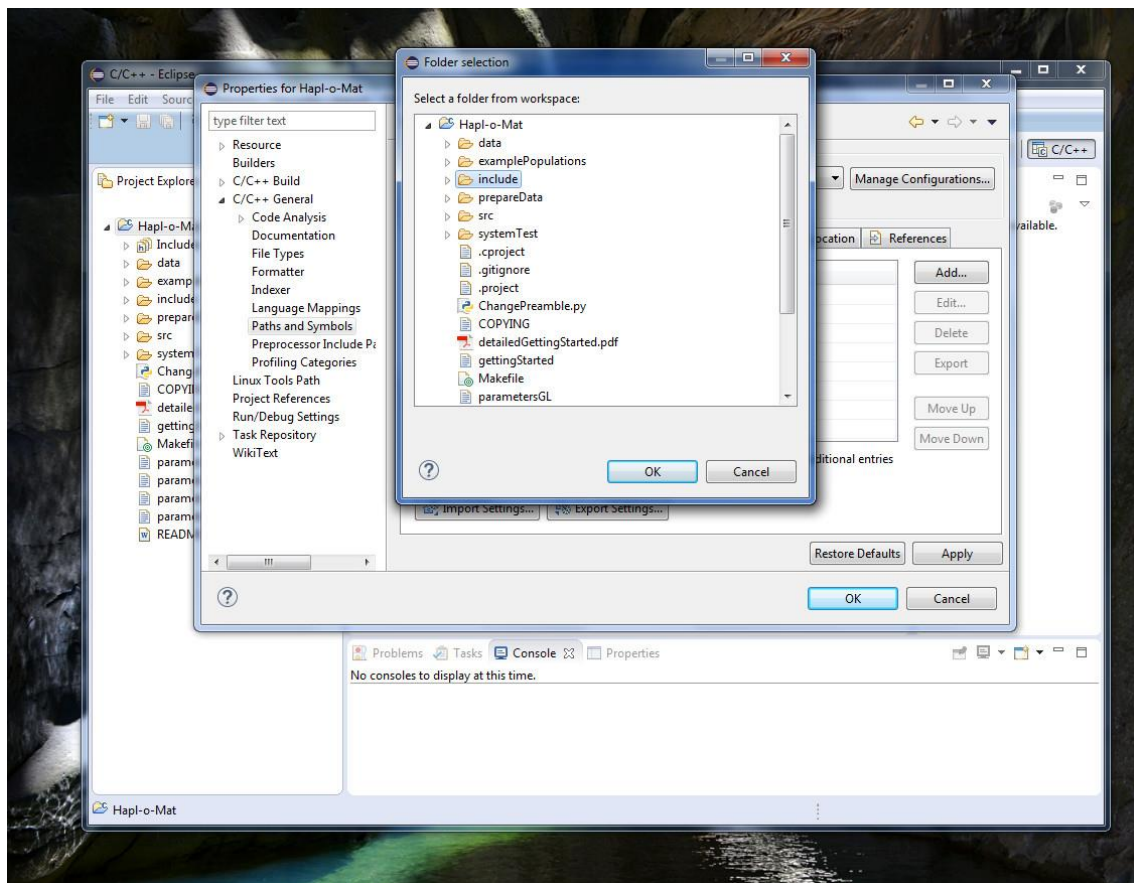
A new window pops up. In the left column choose C/C++ General -> Paths and Symbols. Then, click Add in the right column.



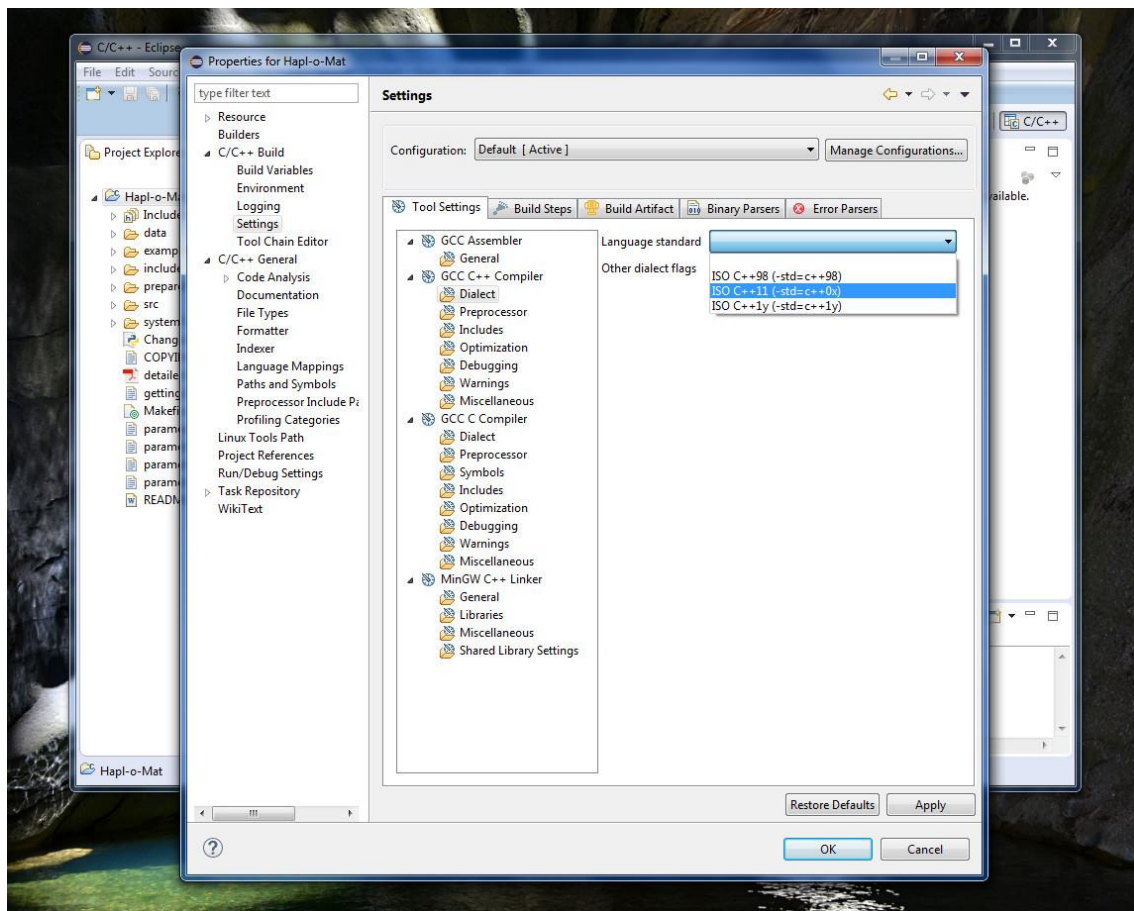
A new window pops up, where you click on Workspace....



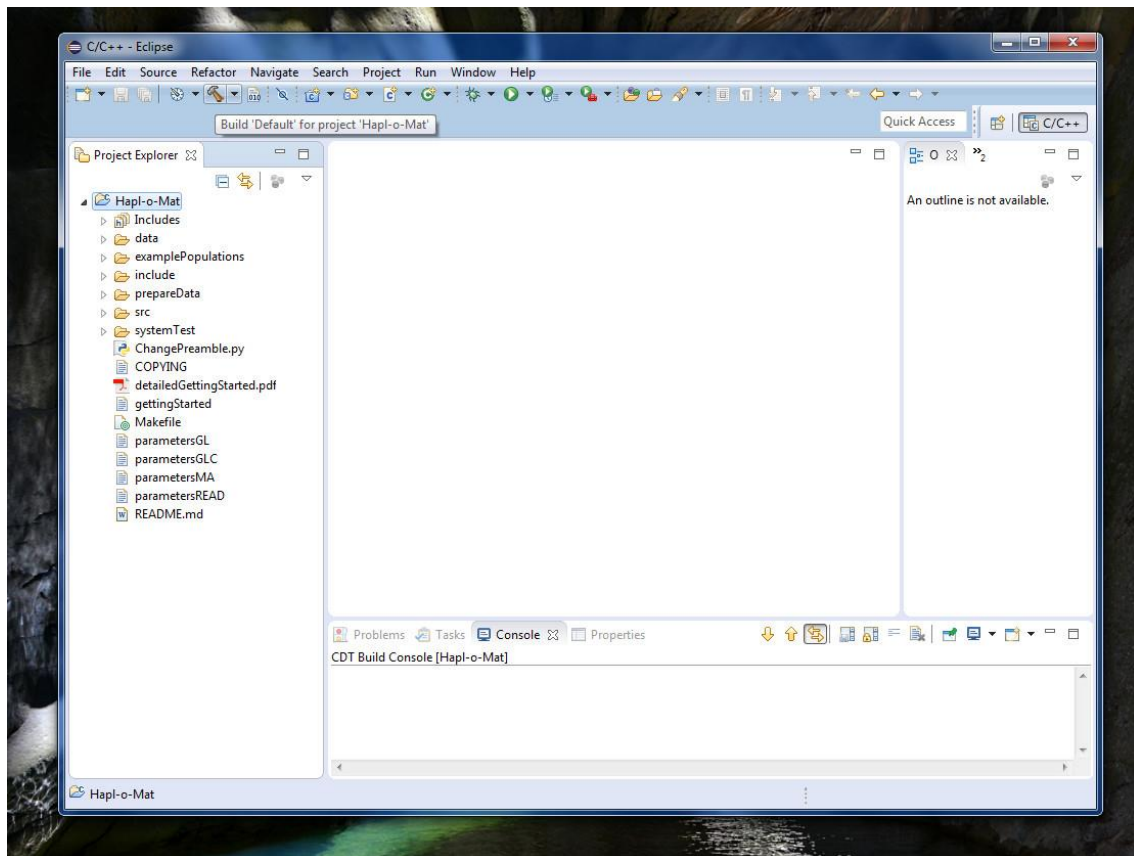
In the following window choose folder include.



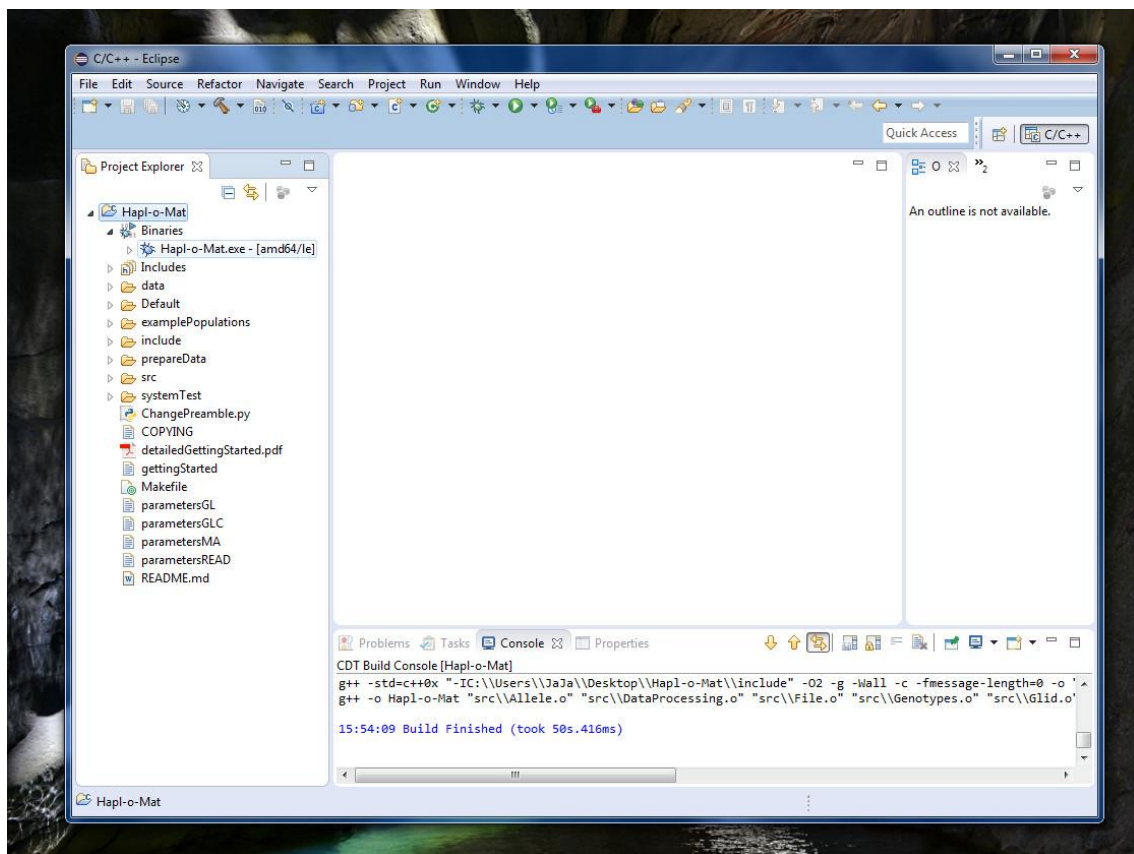
Click two times OK and then choose in the left column C/C++ Build -> Settings. In the right column choose GCC C/C++ Compiler -> Dialect and then choose ISO C++11(-std=c++0x) as Language standard. Click two times OK.



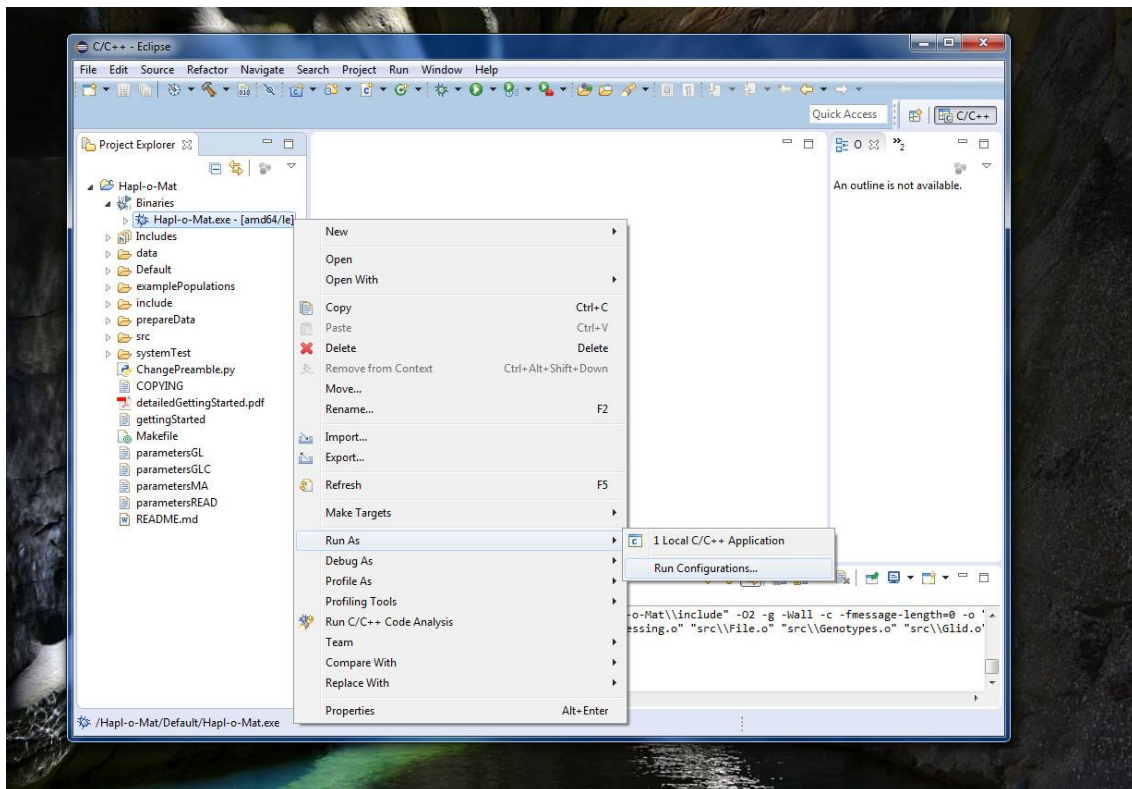
Finally, we compile Hapl-o-Mat by clicking the hammer-symbol in the upper navigation bar.



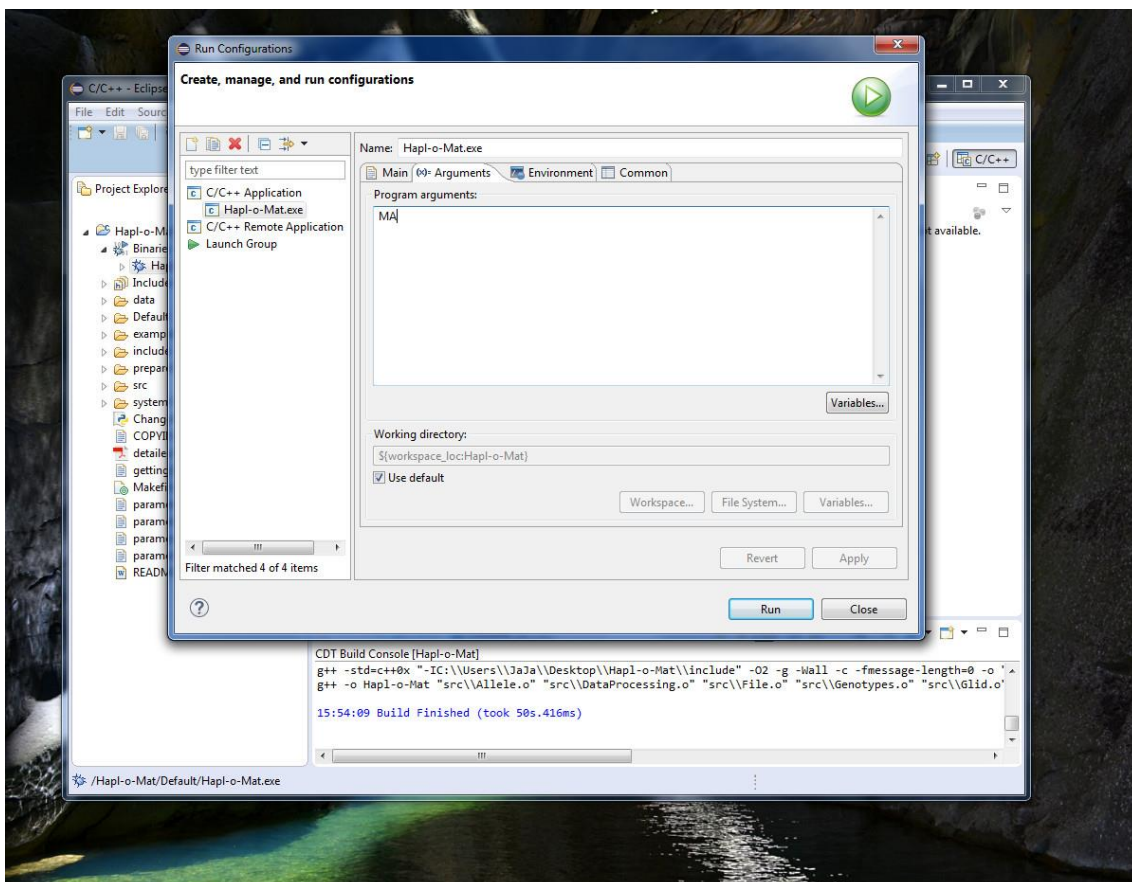
The executable Hapl-o-Mat.exe appeared.



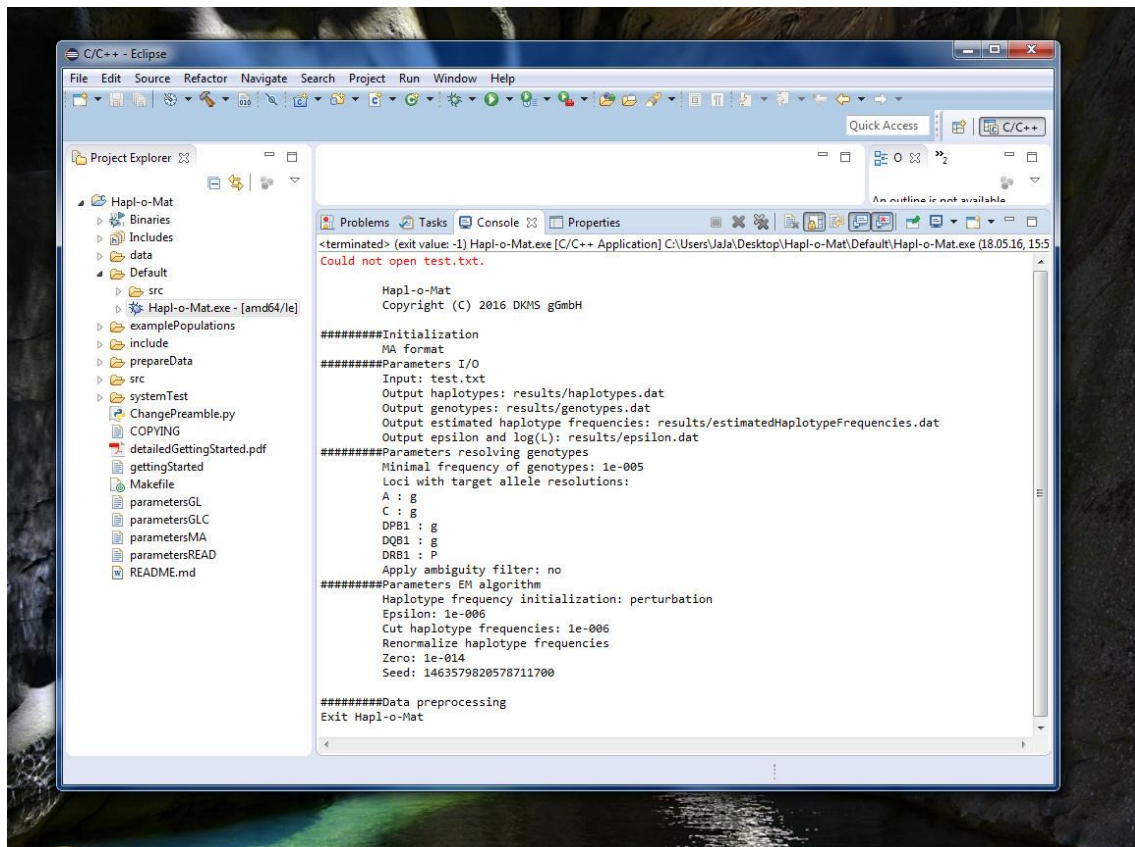
Right-click on it and choose Run As -> Run Configurations....



Click on the tab Arguments and enter MAC into the field Program arguments.



Then click Run. Hapl-o-Mat runs now, but produces an error message, since we did not specify any input files. We tackle this later in the tutorial.



Congratulations, you compiled Hapl-o-Mat under Windows.

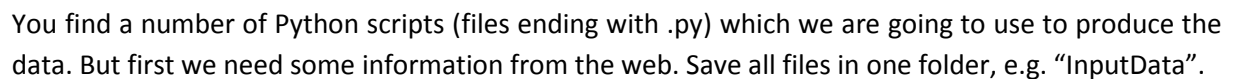
Data Preparation


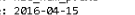
Hapl-o-Mat relies on information on the HLA nomenclature. This information is provided by data files, which we are going to create. As the HLA nomenclature evolves over time, e.g. by finding new alleles or adding new multiple allele codes, it is important to update data from time to time. Hapl-o-Mat relies on the following files, which must be placed in the folder “Hapl-o-Mat/data” (we create the folder “data” later):

File name	Description
AllAllelesExpanded.txt	A list of relevant existing HLA alleles with their enclosed more-digit typing resolutions
AlleleList.txt	If your input data in GLS format includes a missing single-locus genotype, it can be replaced by combining all alleles of the same locus from this file.
Ambiguity.txt	Data basis for the ambiguity filter
LargeG.txt	A list of G-groups with their enclosed alleles in 8-digit resolution
MultipleAlleleCodes.txt	A list of multiple allele codes and their translation to alleles in 4-digit resolution
P.txt	A list of P-groups with their enclosed alleles in 8-digit resolution
Smallg.txt	A list of g-groups with their enclosed alleles in 8-digit resolution

In the following we are going to create these data files.

Check the folder "prepareData".



- 


```

# file: hla_nom_p.txt
# date: 2016-04-15
# version: IPD-INGT/HLA 3.24.0
# origin: http://hla.alleles.org/wmda/hla_nom_p.txt
# author: WHO, Steven G. E. Marsh (steven.marsh@ucl.ac.uk)
A*01:01:01:01/01:01:01:03/01:01:02/01:01:03/01:01:04/01:01:05/01:01:06/01:01:07/01:01:08/01:01:20/01:01:21/01:01:22/01:01:23/01:01:24/01:01:25/01:01:26/01:01:27/01:01:28/01:01:39/01:01:40/01:01:41/01:01:42/01:01:43/01:01:44/01:01:45/01:01:46/01:01:58/01:01:59/01:01:60/01:01:61/01:01:62/01:01:63/01:01:64/01:01:65/01:01:66/01:01:67/01:01:68/01:01:69/01:01:70/01:01:71/01:01:72/01:01:73/01:01:74/01:01:75/01:01:76/01:01:77/01:01:78/01:01:79/01:01:80/01:01:81/01:01:82/01:01:83/01:01:84/01:01:85/01:01:86/01:01:87/01:01:88/01:01:89/01:01:90/01:01:91/01:01:92/01:01:93/01:01:94/01:01:95/01:01:96/01:01:97/01:01:98/01:01:99/01:02:01/02:02:01/02:02:02/02:02:03/02:02:04/02:02:05/02:02:06/02:02:07/02:02:08/02:02:09/02:02:10/02:02:11/02:02:12/02:02:13/02:02:14/02:02:15/02:02:16/02:02:17/02:02:18/02:02:19/02:02:20/02:02:21/02:02:22/02:02:23/02:02:24/02:02:25/02:02:26/02:02:27/02:02:28/02:02:29/02:02:30/02:02:31/02:02:32/02:02:33/02:02:34/02:02:35/02:02:36/02:02:37/02:02:38/02:02:39/02:02:40/02:02:41/02:02:42/02:02:43/02:02:44/02:02:45/02:02:46/02:02:47/02:02:48/02:02:49/02:02:50/02:02:51/02:02:52/02:02:53/02:02:54/02:02:55/02:02:56/02:02:57/02:02:58/02:02:59/02:03:01/03:01:01/03:01:02/03:01:03/03:01:04/03:01:05/03:01:06/03:01:07/03:01:08/03:01:09/03:01:10/03:01:11/03:01:12/03:01:13/03:01:14/03:01:15/03:01:16/03:01:17/03:01:18/03:01:19/03:01:20/03:01:21/03:01:22/03:01:23/03:01:24/03:01:25/03:01:26/03:01:27/03:01:28/03:01:29/03:01:30/03:01:31/03:01:32/03:01:33/03:01:34/03:01:35/03:01:36/03:01:37/03:01:38/03:01:39/03:01:40/03:01:41/03:01:42/03:01:43/03:01:44/03:01:45/03:01:46/03:01:47/03:01:48/03:01:49/03:01:50/03:01:51/03:01:52/03:01:53/03:01:54/03:01:55/03:01:56/03:01:57/03:01:58/03:01:59/03:02:01/03:02:02/03:02:03/03:02:04/03:02:05/03:02:06/03:02:07/03:02:08/03:02:09/03:02:10/03:02:11/03:02:12/03:02:13/03:02:14/03:02:15/03:02:16/03:02:17/03:02:18/03:02:19/03:02:20/03:02:21/03:02:22/03:02:23/03:02:24/03:02:25/03:02:26/03:02:27/03:02:28/03:02:29/03:02:30/03:02:31/03:02:32/03:02:33/03:02:34/03:02:35/03:02:36/03:02:37/03:02:38/03:02:39/03:02:40/03:02:41/03:02:42/03:02:43/03:02:44/03:02:45/03:02:46/03:02:47/03:02:48/03:02:49/03:02:50/03:02:51/03:02:52/03:02:53/03:02:54/03:02:55/03:02:56/03:02:57/03:02:58/03:02:59/03:03:01/03:03:02/03:03:03/03:03:04/03:03:05/03:03:06/03:03:07/03:03:08/03:03:09/03:03:10/03:03:11/03:03:12/03:03:13/03:03:14/03:03:15/03:03:16/03:03:17/03:03:18/03:03:19/03:03:20/03:03:21/03:03:22/03:03:23/03:03:24/03:03:25/03:03:26/03:03:27/03:03:28/03:03:29/03:03:30/03:03:31/03:03:32/03:03:33/03:03:34/03:03:35/03:03:36/03:03:37/03:03:38/03:03:39/03:03:40/03:03:41/03:03:42/03:03:43/03:03:44/03:03:45/03:03:46/03:03:47/03:03:48/03:03:49/03:03:50/03:03:51/03:03:52/03:03:53/03:03:54/03:03:55/03:03:56/03:03:57/03:03:58/03:03:59/03:04:01/04:01:01/04:01:02/04:01:03/04:01:04/04:01:05/04:01:06/04:01:07/04:01:08/04:01:09/04:01:10/04:01:11/04:01:12/04:01:13/04:01:14/04:01:15/04:01:16/04:01:17/04:01:18/04:01:19/04:01:20/04:01:21/04:01:22/04:01:23/04:01:24/04:01:25/04:01:26/04:01:27/04:01:28/04:01:29/04:01:30/04:01:31/04:01:32/04:01:33/04:01:34/04:01:35/04:01:36/04:01:37/04:01:38/04:01:39/04:01:40/04:01:41/04:01:42/04:01:43/04:01:44/04:01:45/04:01:46/04:01:47/04:01:48/04:01:49/04:01:50/04:01:51/04:01:52/04:01:53/04:01:54/04:01:55/04:01:56/04:01:57/04:01:58/04:01:59/04:02:01/04:02:02/04:02:03/04:02:04/04:02:05/04:02:06/04:02:07/04:02:08/04:02:09/04:02:10/04:02:11/04:02:12/04:02:13/04:02:14/04:02:15/04:02:16/04:02:17/04:02:18/04:02:19/04:02:20/04:02:21/04:02:22/04:02:23/04:02:24/04:02:25/04:02:26/04:02:27/04:02:28/04:02:29/04:02:30/04:02:31/04:02:32/04:02:33/04:02:34/04:02:35/04:02:36/04:02:37/04:02:38/04:02:39/04:02:40/04:02:41/04:02:42/04:02:43/04:02:44/04:02:45/04:02:46/04:02:47/04:02:48/04:02:49/04:02:50/04:02:51/04:02:52/04:02:53/04:02:54/04:02:55/04:02:56/04:02:57/04:02:58/04:02:59/04:03:01/04:03:02/04:03:03/04:03:04/04:03:05/04:03:06/04:
```

- 2) Go to the website http://hla.alleles.org/wmda/hla_nom_g.txt and save the file hla_nom_g.txt (same as in 1))

- 3) Go to the website <https://bioinformatics.bethematchclinical.org/HLA-Resources/Allele-Codes/Allele-Code-Lists/Allele-Code-List-in-Alphabetical-Order/>. Click on "Alphabetical Allele Code List (ZIP) (new nomenclature)" and save alpha.v3.zip.

The screenshot shows a web browser window with the URL <https://bioinformatics.bethematchclinical.org/HLA-Resources/Allele-Codes/Allele-Code-Lists/Allele-Code-List-in-Alphabetical-Order/>. The page features a navigation bar with "HLA Resources" (selected), "Search Strategies", "HLA Education", "Policies", and "Contact Us". Below this is a sub-navigation bar with "Haplotype Frequencies", "Allele Codes" (selected), "HaploStats", "HLA Typing", "HML", and "Search Determinants". The main content area is titled "Allele Code List in Alphabetical Order" and includes an example of an allele code (AA 01/02/03/05) and a list of formats. The "Alphabetical Allele Code List (ZIP) (new nomenclature)" link is circled in red. A note states that the self-extracting executable file has been removed as of 10/21/03.

Allele Code List in Alphabetical Order

Example:
AA 01/02/03/05
AB 01/02
AC 01/03

The allele code list in alphabetical order is provided in a variety of formats:

HTML Format (.html)

This format is recommended if you simply want to view the allele code list online.

- [Alphabetical Allele Code List \(HTML\) \(new nomenclature\)](#)
Note: New window will open.
- [Alphabetical Allele Code List \(HTML\) \(old nomenclature\)](#)
Note: New window will open.

Text Format (.txt)

Download a zip compressed file. Once extracted, the text file will be called "alpha.txt."

- [Alphabetical Allele Code List \(ZIP\) \(new nomenclature\)](#)
Note: Extraction requires a data compression program such as WinZip.
- [Alphabetical Allele Code List \(ZIP\) \(old nomenclature\)](#)
Note: Extraction requires a data compression program such as WinZip.

The self-extracting executable file has been removed as of 10/21/03. The allele code lists will no longer be available for download in this format. Please mail new-allelecodes@nmdp.org with any questions or concerns regarding this change.

Extract the archive alpha.v3.zip. Afterwards, you can remove the archive.

- 4) Go to the website <https://www.ebi.ac.uk/ipd/imgt/hla/ambig.html>. Click on "Download Excel" for the wanted release (usually the latest) and save ambiguity_v<>.xls (replace <> by version).

Ambiguous allele combination: x

← → ↻ <https://www.ebi.ac.uk/ipd/imgt/hla/ambig.html>

<https://www.surveymonkey.co.uk/r/F5GHVGF>

Ambiguous Allele Combinations Search Tool (Beta)

This search tool provides an alternative method of viewing the ambiguous combinations currently detailed in the downloads.

STEP 1 - Enter the allele you wish to search for

Query allele:

STEP 2 - Search

[Search Now](#)

Download Ambiguous Allele Combinations files

The IPD-IMGT/HLA Ambiguous Allele Combinations files are available in XML and Microsoft Excel formats for the current release. The PDF versions of the files are no longer being made available from this website because the number of combinations increases rapidly with the release of novel alleles. For this reason, we would encourage our users to use the XML and Microsoft Excel formats where possible. Please note that Microsoft Office 2010 or later is required to view the files.

Older releases are also available in PDF format, as well as XML and Microsoft Excel.

Release	Date	PDF File	Excel File	XML File
3.24.0	2016-04-15	-	Pending	View XML
3.23.0	2016-01-19	-	Download Excel	View XML
3.22.0	2015-10-10	-	Download Excel	View XML
3.21.0	2015-07-06	-	Download Excel	View XML
3.20.0	2015-04-19	-	Download Excel	View XML
3.19.0	2015-01-19	-	Download Excel	View XML
3.18.0	2014-10-10	-	Download Excel	View XML
3.17.0.1	2014-08-21	-	Download Excel	View XML
3.17.0	2014-07-17	-	Download Excel	View XML
3.16.0	2014-04-14	-	Download Excel	View XML

Next we extract the information from the Excel sheet. Open ambiguity_v<>.xls in Excel and save as ambiguity_v<>.xslm to run macros.

ambiguity_v3230.xls [Kompatibilitätsmodus] - Microsoft Excel

Datei Start Einfügen Seitenlayout Formeln Daten Überprüfen Ansicht

[Speichern](#)
[Speichern unter](#)
[Öffnen](#)
[Schließen](#)

Informationen

Zuletzt verwendet

Neu

Drucken

Speichern und Senden

Hilfe

[Optionen](#)

[Beenden](#)

Informationen zu ambiguity_v3230

\\tuesma02\users\cschaefer\Desktop\InputData\ambiguity_v3230.xls

Kompatibilitätsmodus
Einige neue Features sind deaktiviert, um Probleme beim Arbeiten mit früheren Versionen von Office zu verhindern. Die Konvertierung dieser Datei aktiviert diese Features, kann jedoch zu Layoutänderungen führen.

Berechtigungen
Jeder kann diese Arbeitsmappe öffnen und beliebige Teile kopieren und ändern.

Für die Freigabe vorbereiten
Bevor Sie diese Datei freigeben, machen Sie sich bewusst, dass sie Folgendes enthält:

- Dokumenteigenschaften und Name des Autors
- Ausgeblendete Zellen
- Inhalte, die aufgrund des aktuellen Dateityps nicht auf Probleme bei der Barrierefreiheit geprüft werden können

Versionen
Es sind keine früheren Versionen dieser Datei vorhanden.

Eigenschaften

Größe 29,1MB
 Titel Titel hinzufügen
 Kategorien Tag hinzufügen
 Kategorien Kategorie hinzufügen

Verwandte Datumsangaben

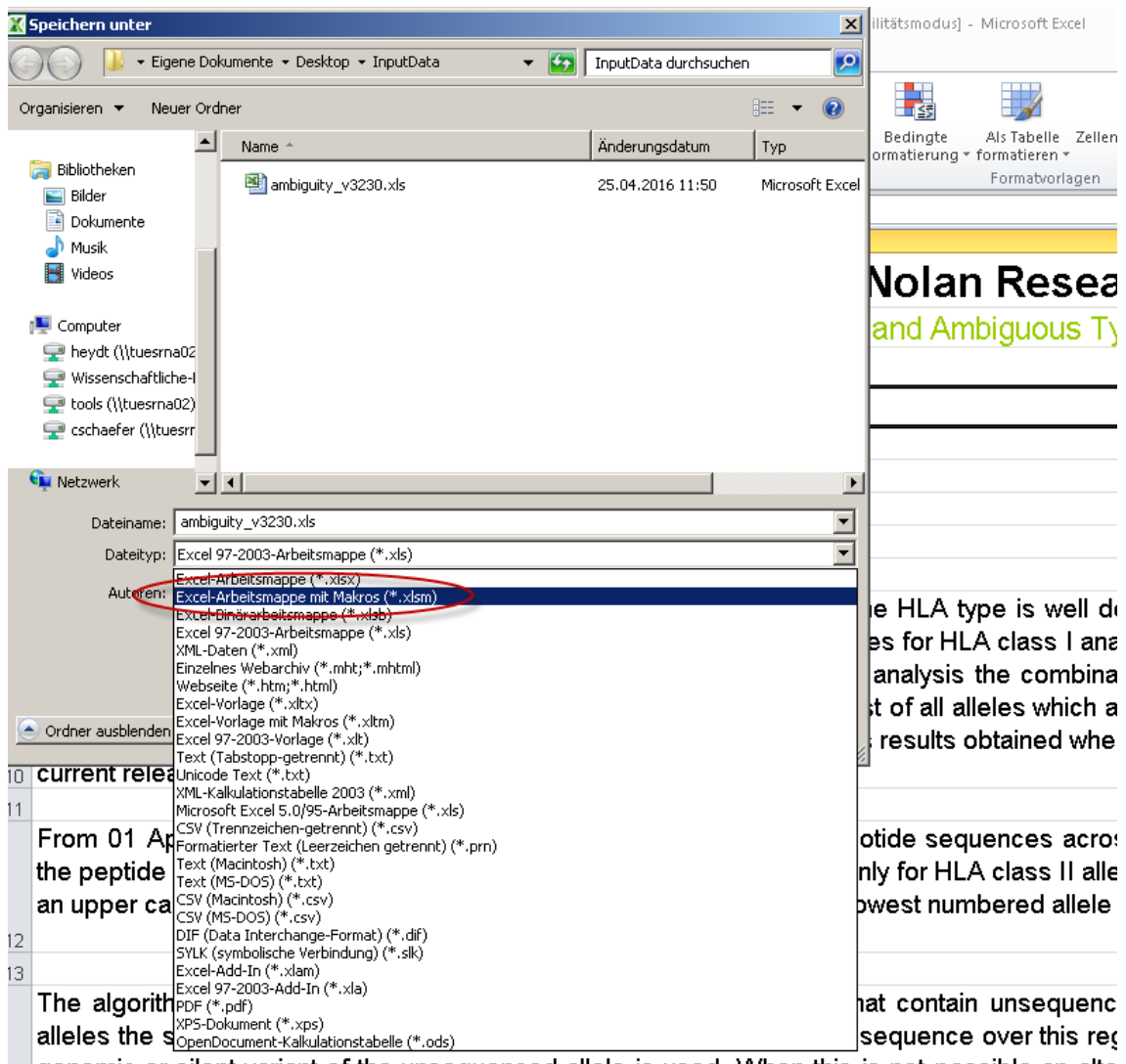
Letzte Änderung 19.01.2016 14:40
 Erstellt 18.01.2016 12:36
 Zuletzt gedruckt Nie

Verwandte Personen

Autor James Robinson
 Autor hinzufügen
 Zuletzt geändert von Anup Raushan Soormally

Verwandte Dokumente

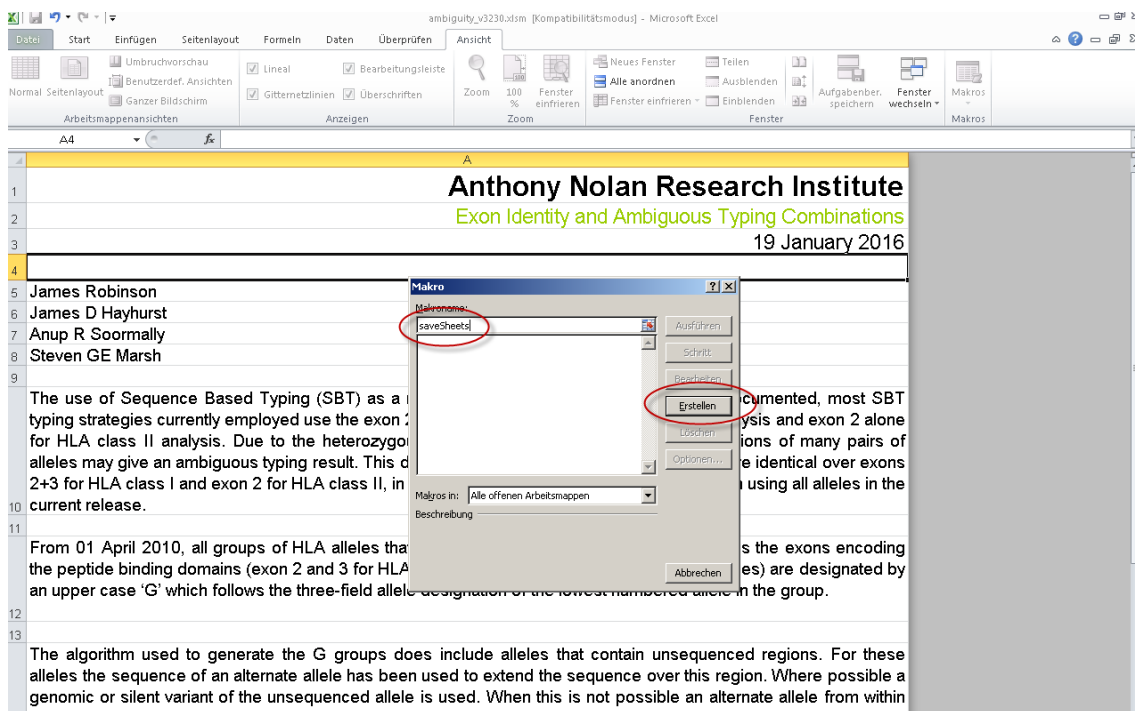
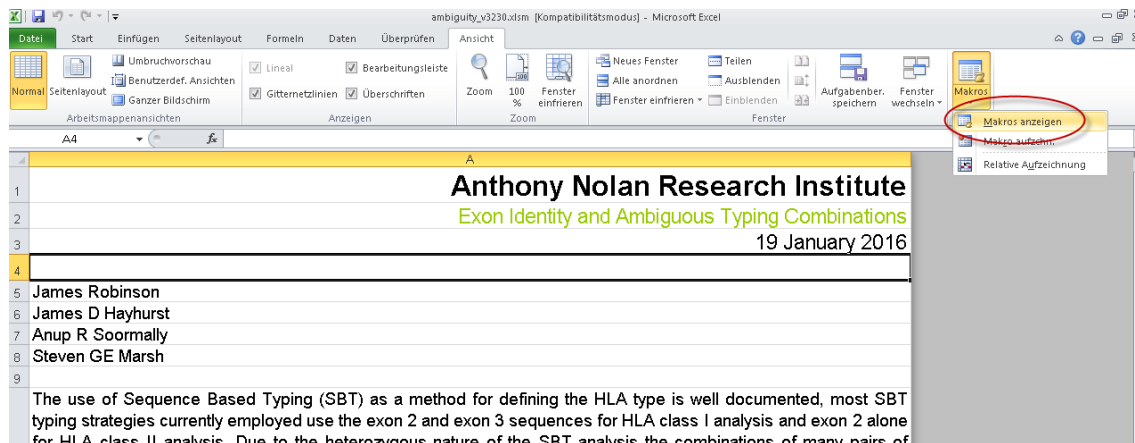
☐ Dateispeicherort öffnen
[Alle Eigenschaften anzeigen](#)

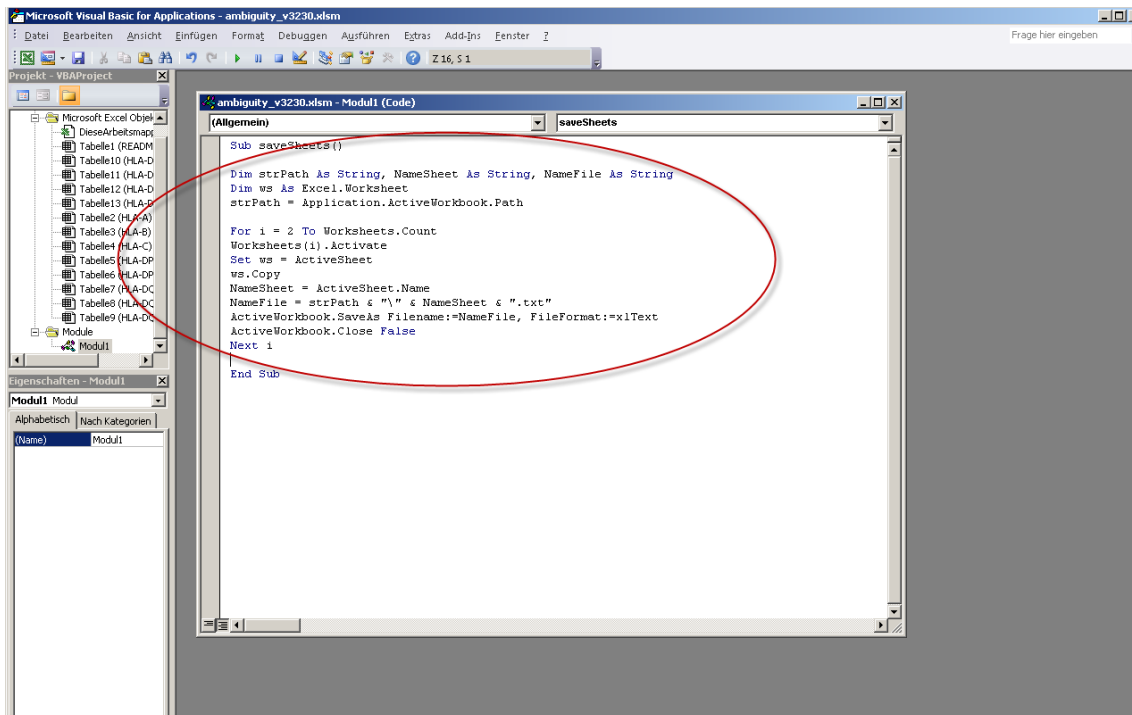


Now insert the following macro, which saves relevant information from the Excelsheets as text files:

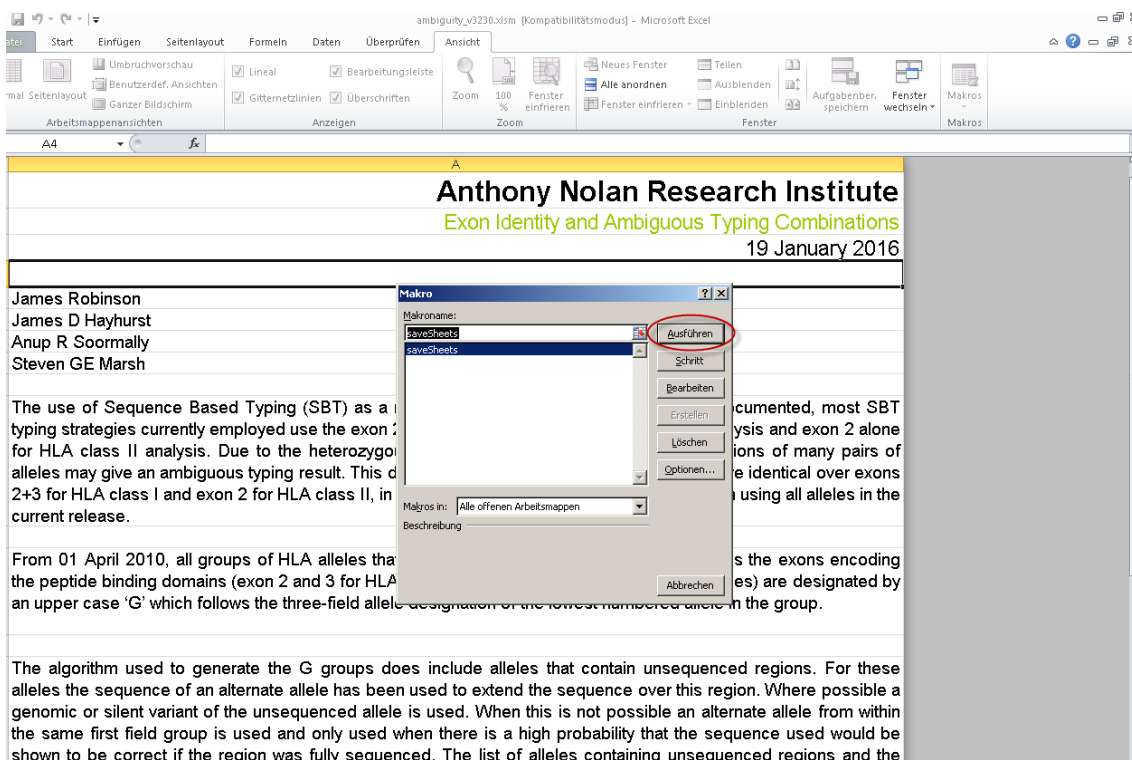
```
Sub saveSheets()
    Dim strPath As String, NameSheet As String, NameFile As String
    Dim ws As Excel.Worksheet
    strPath = Application.ActiveWorkbook.Path

    For i = 2 To Worksheets.Count
        Worksheets(i).Activate
        Set ws = ActiveSheet
        ws.Copy
        NameSheet = ActiveSheet.name
        NameFile = strPath & "\" & NameSheet & ".txt"
        ActiveWorkbook.SaveAs Filename:=NameFile, FileFormat:=xlText
        ActiveWorkbook.Close False
    Next i
End Sub
```





Close the window and execute the macro:



Some new text files should have appeared in your folder „inputData“. Afterwards, you can remove the Excel file.

Build Data for Hapl-o-Mat

Enter the folder InputData and copy all files to the folder “prepareData“. Then enter folder “prepareData“. Next, create the data required for Hapl-o-Mat by running the python scripts in the following order (you can easily get Python here <https://www.python.org/downloads/windows/>):

TransferAlphaToMultipleAlleleCodes.py
 BuildAllAllelesFrom_hla_nom_g.py
 BuildAllAllelesExpanded.py
 BuildP.py
 BuildLargeG.py
 BuildSmallg.py
 BuildAmbiguity.py
 AddGToAmbiguity.py
 AddAllelesMissingIngCode.py

Then create the folder Hapl-o-Mat/data and copy the files

AllAllelesExpanded.txt
 Ambiguity.txt
 LargeG.txt
 MultipleAlleleCodes.txt
 P.txt
 Smallg.txt

to the folder data.

Input Genotype Data

Hapl-o-Mat infers haplotypes from population genotype data. It supports different formats of recording genotype data. To use Hapl-o-Mat your data should be in one of the following data formats:

Data format	Description
MAC	Multiple Allele Codes: ambiguities are encoded by multiple allele codes (MAC). Except for the first line, input files hold an individual's identification number and genotype per line. Genotypes are saved allele by allele without locus name. Identification number and alleles are TAB-separated. The first line of the file is a header file indicating the name of the first column and the loci of the other columns. Same loci must be placed next to each other. For an example refer to "examplePopulations/populationData_a.dat".
GLSC	Genotype List Strings Column-wise: genotypes with or without ambiguities are saved by genotype list strings. Input files hold an individual's identification number and genotype per line. Identification number and single-locus genotypes are TAB-separated. For an example refer to "examplePopulations/populationData_b.dat".
GLS	Genotype List Strings: genotypes with or without ambiguities are saved by genotype list strings (GLS). Population data is saved in two files. The pull-file contains an individual's identification number and a list of integer numbers, GLS-ids, referring to its single-locus genotype. The GLS-ids are separated from the identification number via ";" and from each other via ":". The second file, the glid-file, contains a translation from GLS-ids starting with "1" to actual single-locus genotypes. GLS-id and genotype are separated via ";". A GLS-id of "0" is interpreted as a missing typing at the corresponding locus and does not require a translation in the glid-file. For an example refer to "examplePopulations/populationData_c.pull" and "examplePopulations/populationData_c.glid".

READ	READ: ambiguities are completely resolved and alleles are already translated to the wanted typing resolutions. The input data is of the format as Hapl-o-Mat records processed genotype data. This allows for easily repeating a run without the need to resolve genotype data again.
------	--

When compiling Hapl-o-Mat we assigned MAC as argument to the executable. If you want to run another format, you have to change the argument to the appropriate format abbreviation. Right-click in the Project Explorer and select Run As -> Run Configurations.... Then, click on the tab Arguments and change MAC to the corresponding abbreviation from above.

Parameters

Each input format for genotype data requires a different set of parameters. The parameters are saved in the corresponding files "parametersMAC", "parametersGLSC", "parametersGLS", and "parametersREAD". All input formats have the following parameters in common:

Parameter	Description
FILENAME_HAPLOTYPES	Name of the file which temporarily saves haplotype names
FILENAME_GENOTYPES	Name of the file which saves resolved genotypes
FILENAME_HAPLOTYPEFREQUENCIES	Name of the file which saves haplotypes and estimated haplotype frequencies
FILENAME_EPSILON_LOGL	Name of the file which saves stopping criterion and log-likelihood per iteration
INITIALIZATION_HAPLOTYPEFREQUENCIES	Initialization routine for haplotype frequencies. It takes the following values: <ul style="list-style-type: none"> "equal": All haplotype frequencies are initialized with the same frequency. "numberOccurrence": Haplotype frequencies are initialized according to the initial number of occurrence of the haplotypes. "random": Haplotype frequencies are initialized randomly. "perturbation": Haplotype frequencies are initialized as in numberOccurrence and then randomly modified by a small (<10%) positive or negative offset.
EPSILON	Value for the stopping criterion, i.e. the maximal change between consecutive haplotype frequency estimations is smaller than the assigned value.
CUT_HAPLOTYPEFREQUENCIES	Estimated haplotype frequencies smaller than this value are removed from the output.
RENORMALIZE_HAPLOTYPEFREQUENCIES	Takes values "true" and "false". If "true", normalize estimated haplotype frequencies to sum to one. Within machine precision, this becomes necessary, if estimated haplotypes are removed, e.g. via the option CUT_HAPLOTYPEFREQUENCIES.
SEED	Set the seed of the used pseudo random number

	generator. If set to "0", the seed is initialized by the system time.
--	---

Depending on the input format (indicated in brackets), additional parameters are:

Parameter	Input format	Description
FILENAME_INPUT	MAC, GLSC, READ	The file name of the input population data
FILENAME_PULL	GLS	The file name of the pull-file
FILENAME_GLID	GLS	The file name of the glid-file
LOCI_AND_RESOLUTIONS	MAC, GLS, GLSC	Loci included into analysis and desired typing resolution per locus; The list is separated by "," and contains the locus name followed by ":" and the desired typing resolution, e.g. A:g,B:4d,C:g. Supported typing resolutions and their abbreviations are g-groups (g), P-groups (P), G-groups (G), 2-digit fields (2d), 4-digit fields (4d), 6-digit fields (6d), and 8-digit fields (8d). Alleles are not translated via the option asltIs (applying the ambiguity filter includes an intrinsic translation to G-groups).
LOCIORDER	GLS	Specify the order of loci the individual's GLS-ids correspond to. Loci are separated via ",".
RESOLVE_MISSING_GENOTYPES	GLS	Takes values "true" and "false". If set to true, a missing typing is replaced by a combination of all alleles from AlleleList.txt at the locus. Else, individuals with a missing typing are discarded from analysis.
MINIMAL_FREQUENCY_GENOTYPES	MAC, GLS, GLSC	Genotypes which split into more genotypes than the inverse of this number are discarded from analysis.
DO_AMBIGUITYFILTER	MAC, GLS, GLSC	Takes values "true" and "false". The option "true" activates the ambiguity filter.
EXPAND_LINES_AMBIGUITYFILTER	MAC, GLS, GLSC	Takes values "true" and "false". If set to "true", matching lines with additional genotype pairs in the ambiguity filter are considered.

Whenever specifying a file name including folders, you have to create the folders before running Hapl-o-Mat.

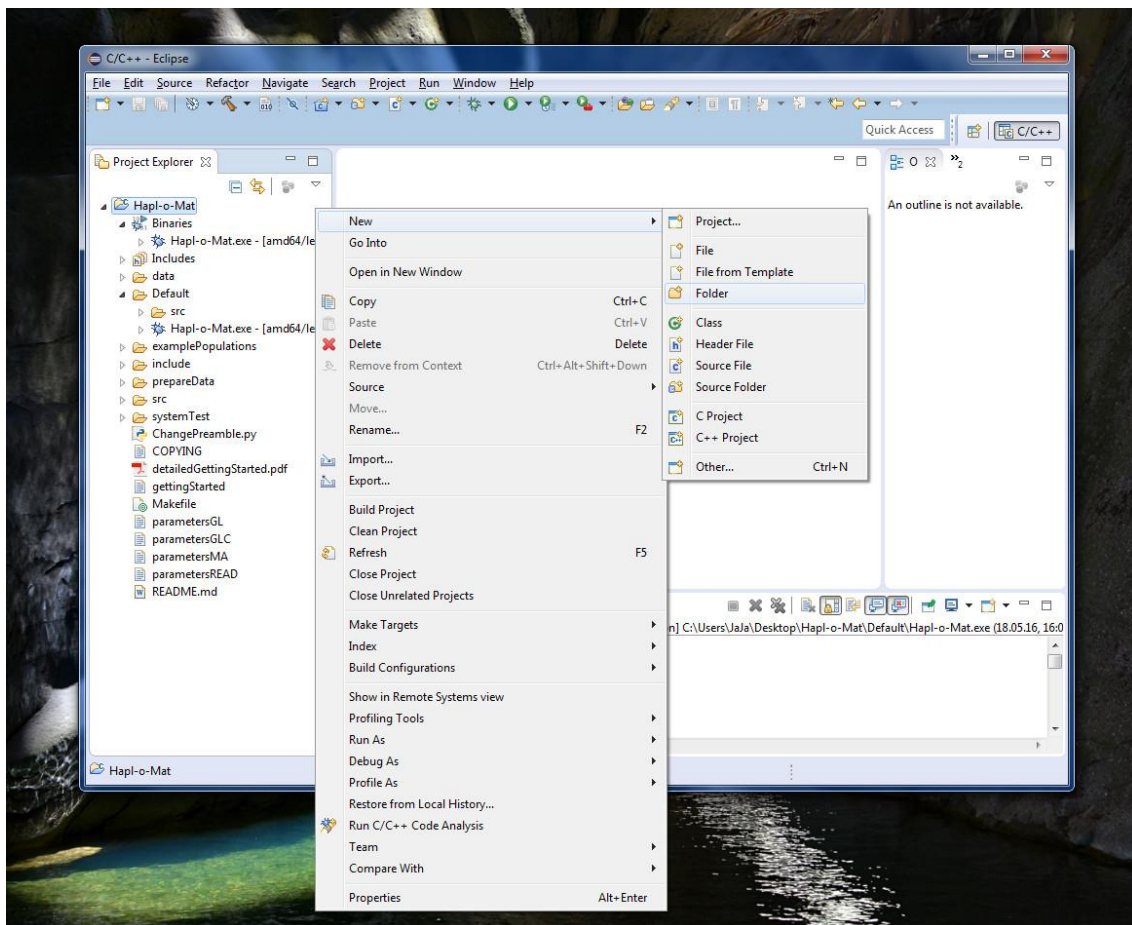
Tutorials

We have everything ready to use Hapl-o-Mat. In the following we estimate haplotype frequencies from some included genotype data recorded in the input format MAC.

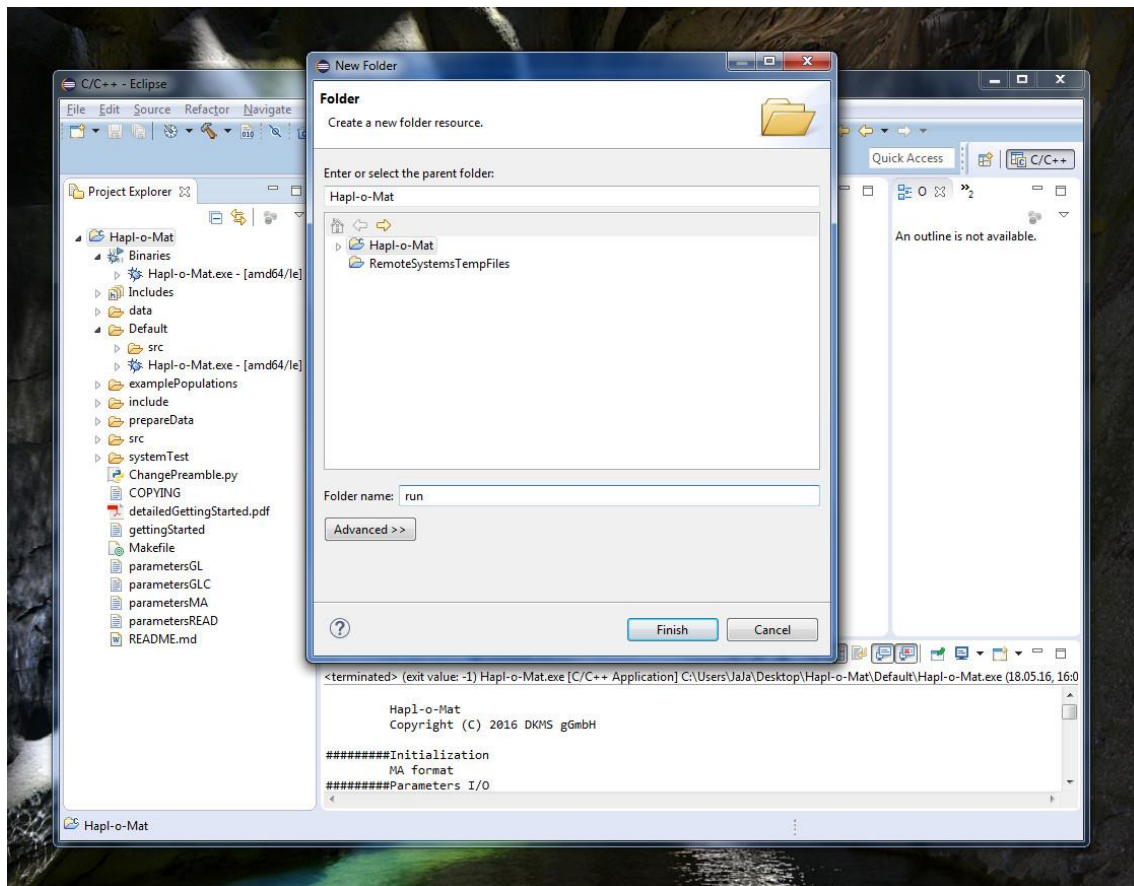
Input Format MAC

You find the relevant population data in "examplePopulations/populationData_a.dat". As ambiguities are recorded as multiple allele codes, the input format is MAC. We are going to infer three locus (A, B, DRB1) haplotypes from this data. Alleles at loci A and B shall be translated to typing resolution g and alleles at locus DRB1 to 4-digits typing resolution.

Create a folder by right-clicking in the Project Explorer and choosing New -> Folder.

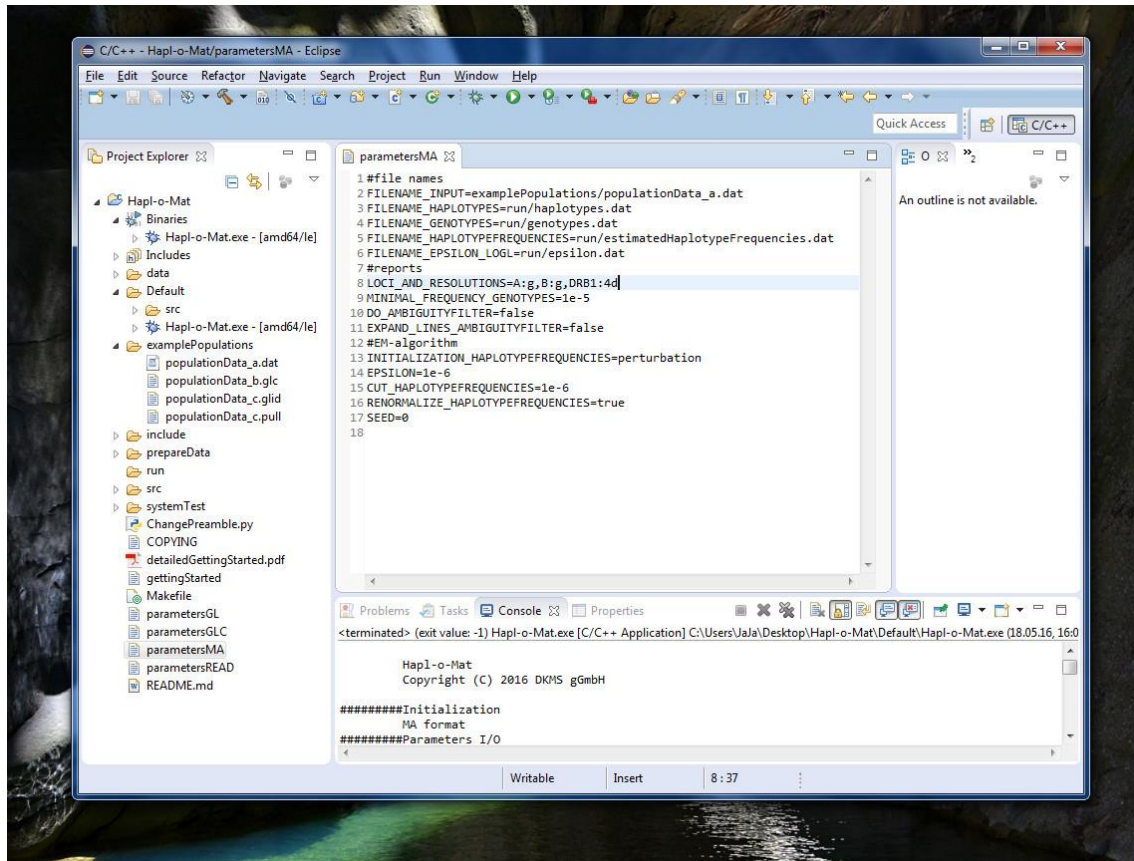


Name the folder run and then click Finish.

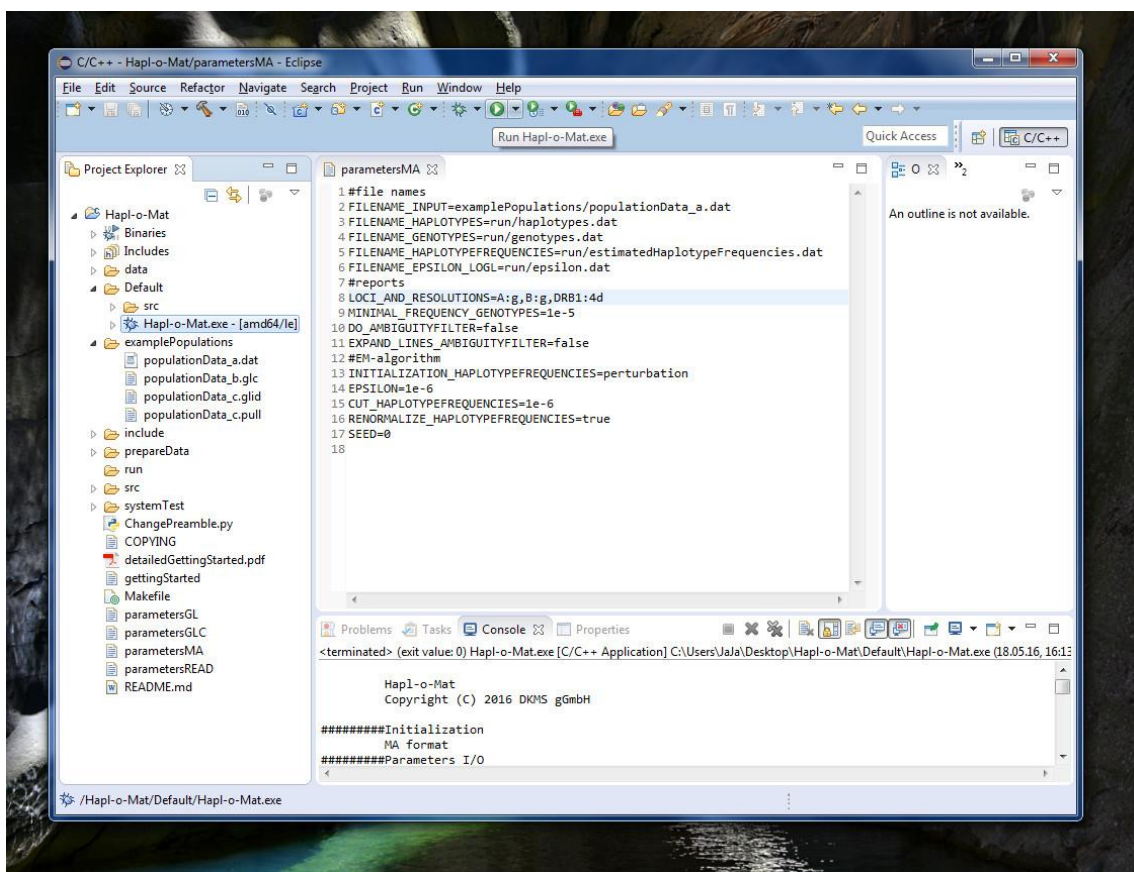


Next, open the parameter file parametersMAC by double-clicking on it. Change its parameters according to:

```
#file names
FILENAME_INPUT=../populationData_a.dat
FILENAME_HAPLOTYPES=run/haplotypes.dat
FILENAME_GENOTYPES=run/genotypes.dat
FILENAME_HAPLOTYPEFREQUENCIES=run/hfs.dat
FILENAME_EPSILON_LOGL=run/epsilon.dat
#reports
LOCI_AND_RESOLUTIONS=A:g,B:g,DRB1:4d
MINIMAL_FREQUENCY_GENOTYPES=1e-5
DO_AMBIGUITYFILTER=false
EXPAND_LINES_AMBIGUITYFILTER=false
#EM-algorithm
INITIALIZATION_HAPLOTYPEFREQUENCIES=perturbation
EPSILON=1e-6
CUT_HAPLOTYPEFREQUENCIES=1e-6
RENORMALIZE_HAPLOTYPEFREQUENCIES=true
SEED=1000
```



Compute haplotype frequencies from the genotype input data by running Hapl-o-Mat. To this end, click on the green Play button in the upper navigation bar.



It produces some output on the screen including your chosen parameters, statistics on the resolved genotype data and the expectation-maximization algorithm, and the run time.

Now let's examine the results produced by Hapl-o-Mat. The results are saved in the folder run (press F5 to update, if they do not appear). We first look into the file with the resolved genotypes, "run/genotypes.dat".

```

1 1 NNN 1 A*26:04+A*29:67*B*07:218+B*54:16*DRB1*11:182+DRB1*13:14
2 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:01*DRB1*12:01+DRB1*14:23
3 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:01*DRB1*12:01+DRB1*14:23
4 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:14*DRB1*12:01+DRB1*14:23
5 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:14*DRB1*12:01+DRB1*14:23
6 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:19*DRB1*12:01+DRB1*14:23
7 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:19*DRB1*12:01+DRB1*14:23
8 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:01*DRB1*12:06+DRB1*14:23
9 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:01*DRB1*12:06+DRB1*14:23
10 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:14*DRB1*12:06+DRB1*14:23
11 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:14*DRB1*12:06+DRB1*14:23
12 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:19*DRB1*12:06+DRB1*14:23
13 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:19*DRB1*12:06+DRB1*14:23
14 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:01*DRB1*12:10+DRB1*14:23
15 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:01*DRB1*12:10+DRB1*14:23
16 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:14*DRB1*12:10+DRB1*14:23
17 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:14*DRB1*12:10+DRB1*14:23
18 2 III 0.098765432098765 A*02:570+A*24:02g*B*13:07N+B*14:19*DRB1*12:10+DRB1*14:23
19 2 III 0.012345679012346 A*02:570+A*24:50*B*13:07N+B*14:19*DRB1*12:10+DRB1*14:23
20 3 NNN 1 A*02:570+A*29:67*B*13:07N+B*35:54*DRB1*13:121+DRB1*14:23
21 4 NIN 0.333333333333333 A*02:77+A*66:10*B*14:01+B*15:154*DRB1*16:30+DRB1*16:30
22 4 NIN 0.333333333333333 A*02:77+A*66:10*B*14:14+B*15:154*DRB1*16:30+DRB1*16:30
23 4 NIN 0.333333333333333 A*02:77+A*66:10*B*14:19+B*15:154*DRB1*16:30+DRB1*16:30
24 5 NNN 1 A*03:217+A*29:36*B*15:01g*B*35:54*DRB1*11:95+DRB1*14:23
25 6 NNN 1 A*02:454+A*03:93*B*15:316+B*35:54*DRB1*04:52+DRB1*08:18
26 7 NNN 1 A*24:02g+A*30:13*B*18:19+B*39:27*DRB1*13:116+DRB1*14:23

```

```

<terminated> (exit value: 0) Hapl-o-Mat.exe [C:\C++ Application] C:\Users\Jala\Desktop\Hapl-o-Mat\Default\Hapl-o-Mat.exe (18.05.16, 17:15)
Different haplotypes: 876

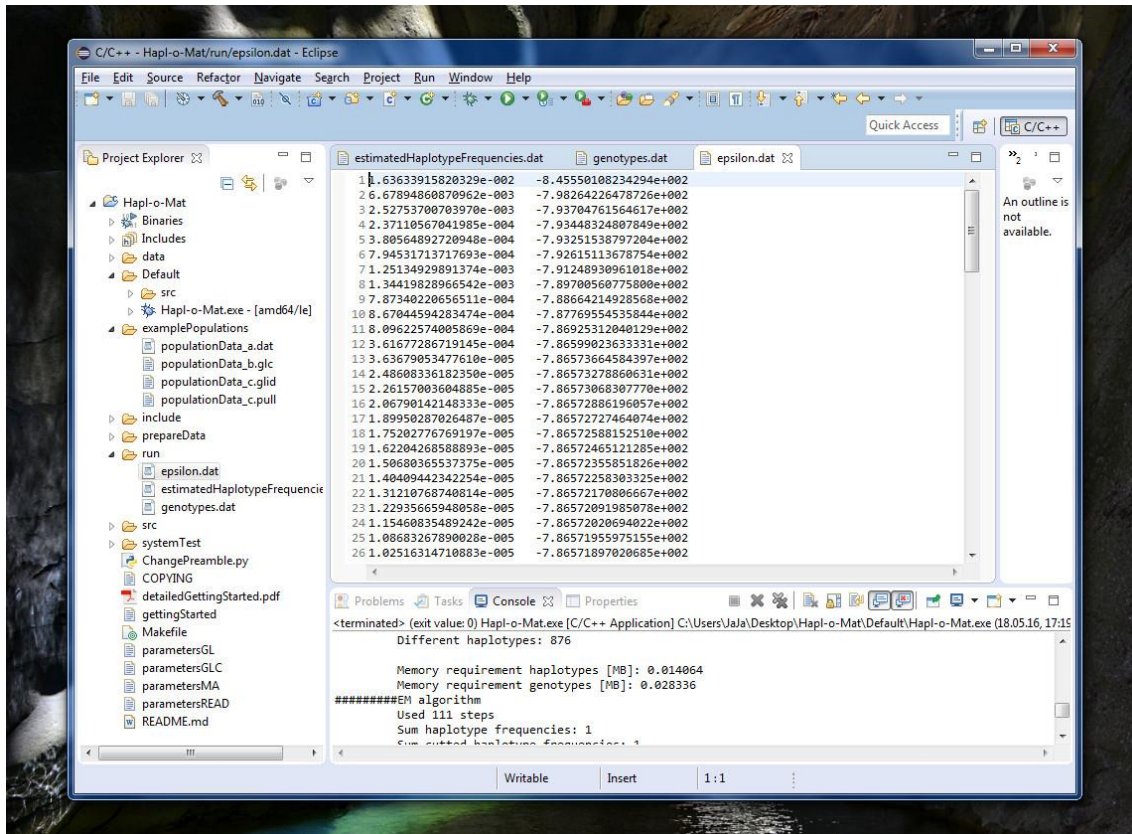
Memory requirement haplotypes [MB]: 0.014064
Memory requirement genotypes [MB]: 0.028336

#####EM algorithm
Used 111 steps
Sum haplotype frequencies: 1
Sum output haplotype frequencies: 1

```

The first column corresponds to the individual's identification number. The second column indicates how ambiguities per single-locus genotypes have been resolved. If no ambiguity occurred or no additional genotypes are formed, the type is N. If an ambiguity occurred and was resolved via building all possible allele combinations, the type is I. Activating the ambiguity filter gives additional types: A, if one matching line in the ambiguity file was found, and M if multiple matching lines were found. The third column gives the frequency of the genotype and the fourth column the genotype itself. The genotype is saved in the GLS format. If an individual's genotype splits into a set of genotypes, each genotype is written to one line starting with the same identification number. The corresponding frequencies become non-integer and sum to one.

The evolution of the stopping criterion and log-likelihood while iterating expectation and maximization steps is written to "run/epsilon.dat". The first column is the stopping criterion and the second one the not normalized log-likelihood.



The inferred haplotypes including estimated frequencies are listed in "run/hfs.dat". Haplotypes are saved in the GLS format. This is the file you were aiming at. It is sorted by descending frequency and already normalized if you activated the corresponding option (we did in this tutorial).

