

BIOINFORMATYKA

1. Wykład wstępny
2. Struktury danych w badaniach bioinformatycznych
3. Bazy danych: projektowanie i struktura
4. Bazy danych: projektowanie i struktura
- 5. Zastosowanie przykładowych programów do analizy danych - równowaga HW**
6. Zastosowanie przykładowych programów do analizy danych
7. Powiązania pomiędzy genami: równ. Hardyego-Weinberga, wsp. rekombinacji
8. Analiza sprzężeń
9. Analiza sprzężeń
10. Analiza asocjacyjna
11. Analiza asocjacyjna
12. Symulacje komputerowe, jackknife, bootstrap
13. Monte Carlo Markov Models
14. Metody klasyfikacyjne
15. Wykład podsumowujący

OMÓWIENIE PROGRAMU



1. Zasady pracy z programem
2. Struktura pliku wsadowego
3. Struktura pliku wynikowego

OMÓWIENIE PROGRAMU

1. ANALIZA GENETYCZNA POPULACJI

- testowanie równowaga HW
- obliczanie współczynnika zaburzenia równowagi HW
- obliczanie dystansu genetycznego między populacjami

2. AUTORZY

- Paul O. Lewis, Dimitri Zaykin

3. PROGRAM NIEKOMERCYJNY

- <http://hydrodictyon.eeb.uconn.edu/people/plewis/software.php>

4. POPRAWNOŚĆ

- autorzy nie gwarantują poprawności działania programu
- w praktyce, do tej pory nie stwierdzono błędów w obliczeniach

OMÓWIENIE PROGRAMU



Home Research People Software Teaching

The software offered below was written by Paul O. Lewis unless otherwise indicated. Most are useful for teaching concepts and are free for downloading. Unfortunately (as one person doing this in my spare time) I can offer very little in the way of support. [email me](#); however, I cannot guarantee that I will have time to address your query (or even answer your email!). I do make requests for help.

Note that these programs are offered AS IS with absolutely NO WARRANTY of any kind.

Data Analysis

[Hickory](#)

Program by Kent E. Holsinger and Paul O. Lewis that provides for Bayesian estimation of F-statistics from both co-dominant marker data.

[GDA](#)

Program by Paul O. Lewis and Dmitri Zaykin designed to accompany the book "Genetic Data Analysis" by Bruce Weir and Michael Donnelly (Sinauer Associates). Computes linkage and Hardy-Weinberg disequilibrium, some genetic distances, and provides methods for hierarchical F-statistics. On **11 January 2008** I changed the download format from self-extracting zip archive to a standard zip archive. Let me know if this causes problems. The new zip file contains an additional example data file (fbi99). This was a request of Bruce Weir to accompany his forthcoming review paper.

[Chris Basten](#) has compiled a command-line-only version of GDA that runs under Mac OS 10.2.8 and 10.3 (Java version can be downloaded [here](#)). After downloading, you should open a terminal window, navigate to the folder, and type "chmod +x gda1.1" to make GDA executable.

PRACA Z PROGRAMEM

GDA - otwieranie pliku wsadowego

The image shows the Genetic Data Analysis (GDA) software interface. The main window has a menu bar with 'File', 'Misc', 'Descr', 'Hierarchy', 'F-stats', 'Dist', 'Diseq', and 'Help'. The 'File' menu is open, showing options: 'Open...', 'Close', 'Editor', 'Log...', 'Import', 'Export', and 'Exit'. An 'Otwórz' (Open) dialog box is overlaid on the main window. The dialog box title is 'Otwórz' and the search path is 'Szyda na 'Gen\Sys\Kg' (H:'. The file list shows folders: 'Access', 'class99_00', 'Excel', 'Internet', 'Msoffice', 'Powerpnt', 'Pulpit', 'Recent', 'Win', 'Word', and 'Skrót do Access'. At the bottom of the dialog, there are fields for 'Nazwa pliku:' (empty) and 'Pliki typu:' (set to 'Nexus (*.nex)'). There are 'Otwórz' and 'Anuluj' buttons.

GDA - definiowanie pliku wynikowego

Genetic Data Analysis

File Misc Descr Hierarchy F-stats Dist Diseq Help

Open...
Close
Editor
Log...
Import ▶
Export ▶
Exit

0.038000	locus-4
1.000000	locus-5
1.000000	locus 1/locus-2
1.000000	
0.038000	
1.000000	
1.000000	
0.293000	
1.000000	
0.133000	
1.000000	
0.038000	

Genetic Data Analysis

File Misc Descr Hierarchy F-stats Dist Diseq Help

Zapisz jako: [?] [X]

Zapisz w: [Folder] [Print] [Grid] [Table]

- Access
- class99_00
- Excel
- Internet
- Msoffice
- Powerpnt
- Pulpit
- Recent
- Win
- Word
- Skrót do Access

Nazwa pliku:

Zapisz jako typ: [v]

Zapisz [Anuluj]

GDA - testowanie równowagi Hardyego-Weinberga

Genetic Data Analysis

File Misc Descr Hierarchy

Genetic Data Analysis
by Paul O. Lewis and Dmitri
32-bit version 1.0 (dl5c) for
Wed Mar 14 13:21:11 2001

Genetic Data Analysis

File Misc Descr Hierarchy F-stats Dist Diseq Help

Disequilibrium Options

Analysis type:

No. shufflings:

Analyze subsets
Up to loci


Only heterozygotes

Measure:
 Fisher Chi-square

Missing data:
 Discard Infer

*** Need: ignoring_gam_prob

GDA - obliczanie współczynnika LD

 **Genetic Data Analysis**
File Misc Descr Hierarchy File

Genetic Data Analysis
by Paul O. Lewis and Dmitri
32-bit version 1.0 (dl5c) fo
Wed Mar 14 13:21:11 2001

Disequilibrium Options ✕

Analysis type: Linkage disequilibrium ▾

Collapse less-frequent alleles

Assume HWE

OK

Estimate

Cancel

GDA - zapisywanie wyników

Genetic Data Analysis

File Misc Descr Hierarchy F-stats Dist Diseq Help

Open...	0.038000	locus-4
Close	1.000000	locus-5
Editor	1.000000	locus 1/locus-2
Log...	1.000000	locus 1/locus-3
Import ▶	1.000000	
Export ▶	0.293000	
Exit	0.132000	
	1.000000	
	0.036000	

Genetic Data Analysis

File Misc Descr Hierarchy F-stats Dist Diseq Help

Clear screen buffer	locus-4
Reset command history	locus-5
Show active loci	locus 1/locus-2
Include/exclude loci	locus 1/locus-3
Monomorphic loci only	locus 1/locus-4
Polymorphic loci only	locus 1/locus-5
Show active populations	locus-2/locus-3
Include/exclude populations	locus-2/locus-4
Preferences...	locus-2/locus-5
	locus-3/locus-4
	locus-3/locus-5
	locus-4/locus-5
	locus 1/locus-2
	locus 1/locus-3
	locus 1/locus-4
	locus 1/locus-5

PLIK WSADOWY

GDA - plik wsadowy

```
#NEXUS
[! Przykladowe dane do analizy ]
begin gdata;
    dimensions nloci=5 npops=2;
    format labels missing=? separator=/;
    locusallelelabels
        1 'locus-1',
        2 'locus-2',
        3 'locus-3',
        4 'locus-4',
        5 'locus-5';

    matrix
    Pop_1:
        _1_ 4/4 4/3 4/3 3/3 4/4
        _2_ 4/4 4/4 4/3 3/3 4/4
        _3_ 4/4 4/4 4/3 4/3 4/4
        _4_ 4/4 4/4 ??/?? 3/3 4/4
        _5_ 4/4 4/4 2/4 3/4 4/4,
    Pop_2:
        _1_ 4/4 4/4 3/3 3/2 4/4
        _2_ 4/4 3/3 4/4 4/3 4/4
        _3_ 4/4 4/3 4/4 4/3 4/4
        _4_ 4/4 4/4 3/3 3/3 4/4 ;

end;
```

GDA - plik wsadowy

#NEXUS

język programowania

[! Przykładowe dane do analizy]

komentarz

begin gdata;

```
dimensions nloci=5 npops=2;
format labels missing=? separator=/;
locusallelelabels
  1 'locus-1',
  2 'locus-2',
  3 'locus-3',
  4 'locus-4',
  5 'locus-5';

matrix
Pop_1:
  _1_ 4/4 4/3 4/3 3/3 4/4
  _2_ 4/4 4/4 4/3 3/3 4/4
  _3_ 4/4 4/4 4/3 4/3 4/4
  _4_ 4/4 4/4 ??? 3/3 4/4
  _5_ 4/4 4/4 2/4 3/4 4/4,
Pop_2:
  _1_ 4/4 4/4 3/3 3/2 4/4
  _2_ 4/4 3/3 4/4 4/3 4/4
  _3_ 4/4 4/3 4/4 4/3 4/4
  _4_ 4/4 4/4 3/3 3/3 4/4 ;
```

dane i instrukcje

end;

UWAGA: każda instrukcja zakończona ;

GDA - plik wsadowy

```
#NEXUS
```

```
[! Przykladowe dane do analizy ]
```

```
begin gadata;
```

instrukcje dla programu

```
dimensions nloci=5 npops=2; liczba loci i populacji
```

```
format labels missing=? separator=/; format
```

```
locusallelelabels
```

```
1 'locus-1',
```

```
2 'locus-2',
```

```
3 'locus-3',
```

```
4 'locus-4',
```

```
5 'locus-5';
```

nazwy poszczególnych loci

```
matrix
```

```
Pop_1:
```

```
_1_ 4/4 4/3 4/3 3/3 4/4  
_2_ 4/4 4/4 4/3 3/3 4/4  
_3_ 4/4 4/4 4/3 4/3 4/4  
_4_ 4/4 4/4 ?/? 3/3 4/4  
_5_ 4/4 4/4 2/4 3/4 4/4,
```

```
Pop_2:
```

```
_1_ 4/4 4/4 3/3 3/2 4/4  
_2_ 4/4 3/3 4/4 4/3 4/4  
_3_ 4/4 4/3 4/4 4/3 4/4  
_4_ 4/4 4/4 3/3 3/3 4/4 ;
```

```
end;
```

GDA - plik wsadowy

#NEXUS

[! Przykladowe dane do analizy]

begin gdata;

dimensions nloci=5 npops=2;

format labels missing=? separator=/;

locusallelelabels

1 'locus-1',

2 'locus-2',

3 'locus-3',

4 'locus-4',

5 'locus-5';

matrix

Pop_1:

1 4/4 4/3 4/3 3/3 4/4

2 4/4 4/4 4/3 3/3 4/4

3 4/4 4/4 4/3 4/3 4/4

4 4/4 4/4 ?/? 3/3 4/4 ←

5 4/4 4/4 2/4 3/4 4/4, ←

Pop_2:

1 4/4 4/4 3/3 3/2 4/4

2 4/4 3/3 4/4 4/3 4/4

3 4/4 4/3 4/4 4/3 4/4

4 4/4 4/4 3/3 3/3 4/4 ; ←

end;

dane

nazwa populacji

genotypy

PLIK WYNIKOWY

1. DANE

- charakterystyka analizowanej próby danych
- opcje dotyczące odczytu danych

2. ANALIZA

- opcje dotyczące analizy danych

3. WYNIKI

- wniosek dotyczący założonych hipotez statystycznych

GDA - plik wynikowy

Opening the NEXUS data file C:\Program Files\gda-1.1.win32\diploid.nex

Plik danych

Data matrix has 6 populations, 5 loci and 44 individuals

Rozmiary zbioru danych

Missing data represented by the symbol ?

Definicja kodu brakujących danych

Different genes at one locus separated by the symbol /

Definicja separatora alleli

Multicharacter allele names.

Wieloznakowe nazwy alleli

Labels provided for individuals.

Pierwsza kolumna danych zawiera nr osobnika

Respecting case for allele names.

Rozróżnia wielkie i małe litery w nazwach alleli

Data matrix not interleaved.

All loci are diploid.

Wszystkie loci są diploidalne

GDA - plik wynikowy

Exact tests for linkage and Hardy-Weinberg disequilibrium

Wyniki testowania równowagi HW

Subsets of loci will be analyzed

Analiza dotyczy również kombinacji loci

Subsets will be comprised of up to 2 loci

Kombinacje składają się z maksymalnie 2 loci

Individuals with missing data will be discarded

Osobniki z brakującymi danymi są pominięte z analizy

Number of runs: 200

Liczba powtórzeń

Measure: Chi-square

Test statystyczny

GDA - plik wynikowy

Population # 1 (Pop 1) of 8 individuals

Wyniki dla próby danych nr 1 złożonej z 8 osobników

Runs Prob Locus combination

Liczba powtórzeń	$\alpha(t)$		kombinacja loci
200	1.000000	H0	locus 1
200	1.000000	H0	locus-2
200	0.155000	H0	locus-3
200	0.485000	H0	locus-4
200	1.000000	H0	locus-5
			Pojedyncze loci
200	1.000000	H0	locus 1/locus-2
200	0.175000	H0	locus 1/locus-3
200	0.455000	H0	locus 1/locus-4
200	1.000000	H0	locus 1/locus-5
200	0.675000	H0	locus-2/locus-3
			Kombinacje loci

GDA - plik wynikowy

Composite disequilibrium measures:

Comparing only most common alleles at each locus
Not assuming Hardy-Weinberg equilibrium

**Wartości wsp. zaburzenia równowagi HW (D)
Obliczone dla najczęstszych alleli w danym locus**

Population # 1 (Pop 1) of 8 individuals

Wyniki dla próby danych nr 1 złożonej z 8 osobników

loci: locus-2/locus-4; alleles: 4/3

Kombinacja loci i nazwy najczęstszych alleli

counts: 2 5 0 1 0 0 0 0 0

Liczebności poszczególnych genotypów

coeffs:	D_a	D_b	D_ab	
estims:	-0.003906	-0.097656	-0.039062	Estymator wsp. D
stdevs:	0.007308	0.053489	0.032682	Odchylenie st. D
chisqu:	0.035556	1.652893	1.904762	Wartość testu
	Locus a	Locus b	pomiędzy	
			loci a i b	



1. **Zasady pracy z programem**
2. **Struktura pliku wsadowego**
3. **Struktura pliku wynikowego**