

Описание пакета программ MANZH-1.0

Пакет компьютерных программ MANZH разработан для комплексного сегрегационного анализа признаков животных. Пакет позволяет анализировать наследование количественных и бинарных признаков по материалу, представленному родословными, полученными при скрещивании представителей одной или двух популяций, притом родословные могут содержать множественные петли. В настоящее время в пакете реализовано несколько моделей наследования признаков: майоргенная модификация смешанной модели и майоргенная модель наследования совместных признаков, характеризующих пару родителей.

Пакет программ написан на языке FORTRAN-77. Основой его послужил ранее разработанный в лаборатории пакет программ МАН-1 (Аксенович, Гинзбург, 1987). В настоящее время пакет реализован для платформы MS-DOS.

Пакет состоит из четырех программ (**recode**, **chstruct**, **peelseq** и **analysis**, см. рис. 1). Три первых являются служебными программами, обрабатывающими структуру анализируемых родословных. Первая из них (**recode**) находит в исходных данных отдельные родословные, делает перекодировку и сортировку, т.е. преобразует данные в формат, удобный для дальнейшего анализа. Вторая (**chstruct**) проверяет, присутствуют ли в родословной петли и, при необходимости, преобразует структуру родословной. Третья (**peelseq**) находит последовательность свертки родословной.

Последняя программа (**analysis**) позволяет проводить комплексный сегрегационный анализ: она производит подсчет и максимизацию функции правдоподобия.

Подсчет ФП, $L(\vec{x}|\mathcal{S}, \vec{\theta})$, требует знания вектора фенотипов членов родословной \vec{x} , структуры родословной \mathcal{S} и вектора параметров $\vec{\theta}$. Кроме того, для применения эффективного алгоритма вычисления ФП необходимо знать последовательность свертки \vec{P}_s .

Вектор фенотипов членов родословной хранится в файле **tra.tra**, структура родословной и последовательность свертки хранятся в файлах **iway.iwa** и **iway.iyw**, а набор параметров, ограничения, наложенные на них и их начальные значения — в файле **model.inp**. Кроме того, файл **lparam.inp** содержит указания о способе подсчета ФП: применяется ли точный метод или приближительный, если приближительный, то какой именно.

Программы обработки структуры родословной

Программа **recode**

Исходные данные о структуре родословной представлены набором записей, по одной на каждую особь. Запись, характеризующая некоторую особь, состоит из пяти целочисленных полей, это: уникальный номер особи, номера ее матери и отца (если не известны, ставится 0), популяционная принадлежность особи (1, 2 или 0 — если для особи известны родители) и пол (1 — женский, 2 — мужской). Последовательность записей значения не имеет. Примером исходного файла данных может служить **pedigree.prn**:

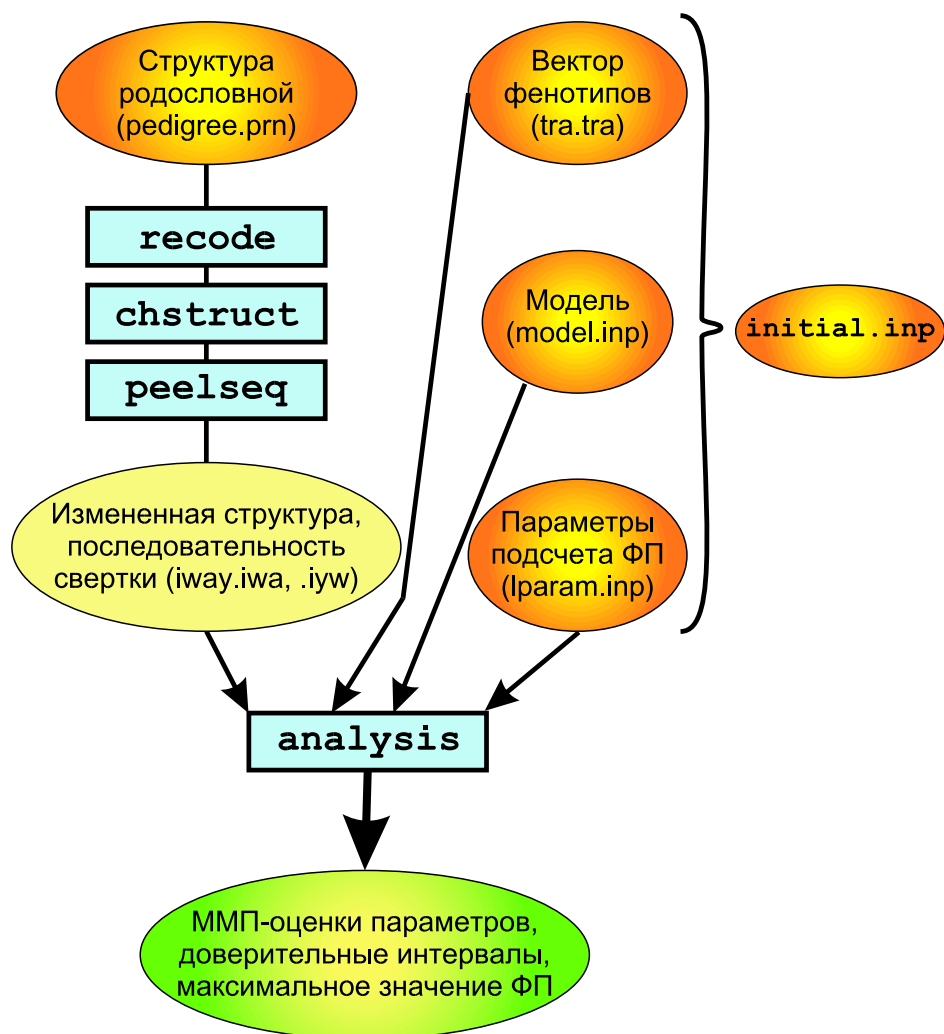


Рис. 1: Структура пакета программ MANZH-1.0. Прямоугольники — программные модули, овалы — блоки данных и параметров.

```

INI_ID MATHER_ID FATHER_ID POPULATION GENDER
9
11 0 0 1 2
12 0 0 2 1
112 12 11 0 2
212 12 11 0 2
312 12 11 0 2
412 12 11 0 1
21 0 0 1 2
113 412 21 0 1
114 113 11 0 1

```

Программа **recode** сортирует эти записи, находит отдельные родословные, перекодирует номера особей. Имена входного и выходного файлов указаны в управляющем файле **rewdata.par**:

```

#
# Input and output data file names PRINTED
# in one line separated by space
#
#
pedigree.prn    pedigree.1st

```

После запуска программы, она считывает данные из файла **pedigree.prn**, обрабатывает их, и записывает результаты в файл **pedigree.1st**:

```

9      2
family number
individual number
father's number
mother's number
race
sex
initial ID
1      1      0      0      1      2      11
1      2      0      0      2      1      12
1      3      2      1      0      2      112
1      4      2      1      0      2      212
1      5      2      1      0      2      312
1      6      2      1      0      1      412
1      7      0      0      1      2      21
1      8      6      7      0      1      113
1      9      8      1      0      1      114

```

Видно, что после обработки изменяется нумерация особей и появляется два дополнительных поля — это номер родословной и исходный номер особи.

Программа chstruct

Пакет программ MANZH позволяет проводить анализ наследования по родословным с множественными петлями. В пакете реализовано два алгоритма, позволяющие проводить такой анализ — это алгоритм Отта, позволяющий точный подсчет ФП, и несколько вариантов алгоритма Ванга-Стрикера, позволяющего находить приближенное значение ФП. Оба

алгоритма требуют найти набор особей, дублирование которых позволяет превратить родословную в граф-дерево. Именно эту задачу решает программа **chstruct**. После запуска, программа требует вести имя файла, хранящего информацию о структуре родословной, и способ дублирования:

```
E:\MANG-1.0\STEP2>CHSTRUCT.EXE
*****
* INPUT NAME OF FILE WITH SAMPLE DATA: pedigree.1st

CHESK of errors in pasport data :
NOT ERRORS in pasport data

WHAT TYPE OF LOOPS BREAKS DO YOU WISH TO USE?
  1 -- STRICKER; 2 -- MULTIPLICATION
  3 -- CIL-STRICKER; 4 -- CIL-MULTIPLICATION
  5 -- OPTIMIZE FOR EXACT (CIL)
  6 -- WANG-STRICKER CUT-EXTENSION
1
      8
      1      8
BREAKAGE: NFA,NMA      8      1
Stop - Program terminated.
```

Алгоритмы 1–6 делятся на 2 группы — одна предполагает разрыв всех петель в родословной, а другая, маркированная аббревиатурой CIL, требует применять дополнительную свертку коротких инбредных петель (документ "introd-CSA-maia", глава 3) при подсчете ФП. К сожалению, в настоящий момент эта свертка реализована не для всех классов моделей и мы не рекомендуем использование алгоритмов 3, 4, и 5. Алгоритм 2 оптимален, если в дальнейшем планируется подсчитывать точную ФП с помощью алгоритма Отта. Алгоритмы 1 и 6 являются вариантами алгоритма Ванга-Стрикера — первый из них реализует базовый вариант алгоритма, а второй позволяет дополнительное расширение, как описано в документе introd-CSA-maia, глава 3, вариант 3.

Результатом работы программы **chstruct** является модифицированная родословная записанная в файл **pedigree**:

10	2					
1	1	0	0	1	2	11
1	2	0	0	2	1	12
1	3	2	1	0	2	112
1	4	2	1	0	2	212
1	5	2	1	0	2	312
1	6	2	1	0	1	412
1	7	0	0	1	2	21
1	8	6	7	0	1	113
1	9	10	1	0	1	114
1	10	0	0	100	1	113

Видно, что в результате работы была дублирована особь с исходным номером 113. Копия помечена "породной" принадлежностью 100.

Программа `peelseq`

Программа `peelseq` определяет последовательность свертки и переписывает данные в формате, наиболее удобном для дальнейшего анализа. Исходные файлы данных указаны в файле `input1.dat`:

```
pedigree
race
```

Результатом работы программы являются два файла — `iway.iwa` и `iway.iyw`. Далее, эти файлы используются программой `analysis`.

Программа комплексного сегрегационного анализа `analysis`

В настоящее время существует три варианта программы, реализующих три модели наследования — это майоргенная модель наследования признака в гибридной популяции (количественных QT и бинарных BIN-QT, см. раздел 2 документа `introd-CSA-maia`, глава 2) и майоргенная модель наследования совместного признака, характеризующего пару родителей (COMTRA, см. раздел 3 документа `introd-CSA-maia`, глава 3).

Гибкое применение программы `analysis`, возможность модификации ее для требований конкретного анализа требует знания идеологии, реализованной в программе.

Исходный код программы делится на две части — это ядро программы, которое не зависит от модели наследования, и надстройка — программные модули, в которых реализована сама модель. Компиляция ядра и конкретной настройки дает программу для анализа наследования конкретного признака. Эта идеология сходна с идеологией пакета программ `MENDEL`.

Кроме файлов, определяющих структуру родословной и последовательность свертки (`iway.iwa` и `iway.iyw`) для подсчета ФП необходимо определить модель (файл `model.inp`) и вектор фенотипов (файл `tra.tra`).

Файл, описывающий модель, содержит несколько служебных записей, а так же по одной записи на каждый из параметров модели.

Наиболее часто встречаются следующие параметры:

Идентификатор параметра	Описание
BETA(TR COV1	коэффициент регрессии признака на дополнительную переменную 1
BETA(TR COV2	коэффициент регрессии признака на дополнительную переменную 2
BETA(TR COV3	коэффициент регрессии признака на дополнительную переменную 3
BETA(TR COV4	коэффициент регрессии признака на дополнительную переменную 4
PR(A2,P0)=Q1	частота аллеля A_2 в популяции 0
PR(A2,P1)=Q2	частота аллеля A_2 в популяции 1
Z0(POL.LEV.0	среднее значение признака при условии майоргенотипа A_2A_2 в популяции 0
Z1(POL.LEV.1	среднее значение признака при условии майоргенотипа A_2A_2 в популяции 1
MG(A1A1,S1)	прибавка к среднему при условии майоргенотипа A_1A_1
MG(A1A2,S1)	прибавка к среднему при условии майоргенотипа A_1A_2
NJU(HETEROZ)	максимальное изменение значения признака за счет гетерозиса, η
SIGMA(S1,P0)	корень из внемайоргенной дисперсии признака в популяции 0
SIGMA(S1,P1)	корень из внемайоргенной дисперсии признака в популяции 1
GAMMA(S1)	добавочная полигенная дисперсия γ
PR(A1/A1A1)	переходная вероятность 1
PR(A1/A1A2)	переходная вероятность 2
PR(A1/A2A2)	переходная вероятность 3
THRESHOLD	

Каждый параметр характеризуется записью из 8-ми полей:

```
C QUANTITATIVE TRAIT LIABILITY MODEL; ELSTON-AXENOVICH, 1971,1999
C
C
  25          % число параметров
PCONST       1.0000E+01 % служебная константа
BETA(TR|COV1 1 1 1-0.10000E+00 0.05000E-00-0.10000e+04 0.10000e+04
BETA(TR|COV2 2 0 0 0.00000E+00 0.10000E-00-0.10000e+04 0.10000e+04

...

PR(A1/A1A2) 23 1 0 0.50000E+00 0.10000E-00 0.00000e+00 1.00000e+00
PR(A1/A2A2) 24 1 0 0.00000E+00 0.10000E-00 0.00000e+00 1.00000e+00
THRESHOLD   25 1 0 0.50000E+00 0.10000E-00-1.00000e+02 1.00000e+02
|           | | | | |           |           |           |
|           | | | | |           |           |           \___ правая граница
|           | | | | |           |           \___ левая граница
|           | | | | |           \___ начальный шаг максимизации
|           | | | \___ начальное значение
|           | | \___ участие в максимизации (IMAX)
|           | \___ участие в модели (IMOD)
|           \___ порядковый номер параметра (NP)
\___ идентификатор параметра (ID)
```

Порядковый номер параметра NP может принимать два типа значений: это действительно порядковы номер, либо номер, который меньше порядкового. В последнем случае значение параметра автоматически приравнивается значению параметра с этим номером.

Поле IMAX может принимать два значения — 1, если параметр максимизируется, или 0, если параметр зафиксирован в начальном значении.

Участие в модели (0 — не участвует, 1 — участвует) имеет смысл для коэффициентов регрессии (при IMOD=0 они исключаются из модели). Для остальных параметров значение IMOD в основном варианте пакета должно быть равно 1.

Файл, содержащий вектор фенотипов членов родословной, организован следующим образом:

```
% /--- число особей в родословной
% | /--- число измеренных признаков
% | |
  556 5 NIP_ASSYM period NIP_SUM EYE_ASSYM HTH_FULL
% /--- исходный номер особи
% | /--- число измерений признака
% | |
  26 1
% /--- доля крови популяции 1 у матери особи
% | /--- доля крови популяции 1 у отца особи
% | |
.0000E+00 .1000E+01
.0000E+00 % значение признака
.0000E+00 % значение дополнительной переменной 1
.0000E+00 % значение дополнительной переменной 2
.1200E+02 % значение дополнительной переменной 3
.0000E+00 % значение дополнительной переменной 4

...

```

Кроме того, способ подсчета ФП задается файлом `lparam.inp`:

```
C 0/1 = SCREENING/AAF      % 0 --- прямой подсчет ФП, 1 --- условный
C 0/1 = G-COPYING OFF/ON % 0 --- приблизительный подсчет, 1 --- алгоритм Отта
C 0/1 = ERRMAL OFF/ON % 0 --- не вычислять 95% CI, 1 --- вычислять
C 0/1 = NOT/FIND GENOTYPE PROBABILITIES % 1 --- нахождение вероятностей генотипов
% членов родословной
0
1
0
0
```

Определение описанных 5-ти файлов позволяет начать подсчет и максимизацию ФП. Результаты анализа записываются в файл, по структуре своей сходный с файлом `model.inp`, в котором вместо начальных значений параметров стоят ММП-оценки, и, кроме того, приводится максимальное значение ФП.