Multiple Trait Analysis

observations on multiple traits
one or more traits in multiple environments

Does QTL have pleiotropic effects on multiple traits?
Does QTL show genotype-environment interaction?
What is genetic correlation between different traits?
Is correlation due to pleiotropy or linkage? Where?

view multiple traits as multivariate vector
Falconer (1952); Jiang Zeng (1995)

statistical models and likelihood analyses
hypothesis tests of QTL effects
pleiotropy vs. close linkage
QTL by environment interaction

Statistical Models and Likelihood Analyses

CIM model for multiple traits

sample of \(n\) individuals from a \(F_2\) population
additive effects only (for now)
observe \(m\) quantitative traits
CIM scan for QTL on a marker interval \((M_i, M_{i+1})\)

\[
y_{j1} = b_{01} + a_1^* x_j^* + \sum b_{1l} x_{jl} + e_{j1}
y_{j2} = b_{02} + a_2^* x_j^* + \sum b_{2l} x_{jl} + e_{j2}
\vdots \nonumber
y_{jm} = b_{0m} + a_{m}^* x_j^* + \sum b_{ml} x_{jl} + e_{jm}
\]

\(y_{jk}\): phenotype of \(k\)th trait on individual \(j\)
\(b_{0k}\): mean effect (reference) for trait \(k\)
\(a_{k}^*\): additive effect of putative QTL on trait \(k\)
\(x_{jl}\): genotype at marker \(l\)
\(b_{lk}\): marker regression coefficients
\(e_{jk}\): residual effect on trait \(k\) for individual \(j\)

assumptions on residual “error” effects \(e_{jk}\)

errors correlated among traits within individuals
covariance: \(\text{Cov}(e_{jk}, e_{jl}) = \sigma_{kl} = \rho_k \sigma_k \sigma_l\)
independent individuals: \(\text{Cov}(e_{jk}, e_{ik}) = 0\)

variance-covariance matrix
errors multivariate normal among individuals
mean zero
general covariance matrix \((\sigma_{kl} = \sigma_{lk})\)

\[
V = \begin{pmatrix}
\sigma^2_1 & \sigma_{12} & \cdots & \sigma_{1m} \\
\sigma_{21} & \sigma^2_2 & \cdots & \sigma_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{m1} & \sigma_{m2} & \cdots & \sigma^2_m
\end{pmatrix}
\]

vector notation

\[
y_j = x_j^* a^* + x_j^T B + e_j
\]

\(y_j\): vector of phenotypes \(y_{jk}\)
\(a^*\): vector of QTL effects \(a_{k}^*\)
\(x_j\): vector of \(1\) and marker data \(x_{jk}\)
\(B\): \((n_p + 1) \times m\) matrix of cofactor effects
\(= \text{(reference } b_{0k}\text{ and cofactors } b_{lk}\text{)}\)
\(e_j\): vector of errors \(e_{jk}\)
\(\text{Cov}(e_{j}) = V\text{ covariance matrix}\)

matrix notation

\[
Y = x^* a^* + XB + E
\]

\(Y\): \(n \times m\) matrix of \(y_{jk}\) (row \(j = y_j\))
\(x^*\): \(n\) vector of \(x_{j}^*\)
\(X\): \(n \times (n_p + 1)\) marker matrix (column \(j = x_j\))
\(E\): \(n \times m\) error matrix of \(e_{jk}\) (row \(j = e_j\))

choice of background markers?
additive and dominance effects?
same issues for selecting cofactors as ordinary CIM
likelihood analysis

\[ Y = x^T a^* + XB + E \]
\[ y_j = x_j^* a^* + x_j^T B + e_j \]

\[ Cov(e_j) = V \]

However, we do not know \( x^* = \{x_j^*\} \) mixture model with multivariate normal

\[ L_1 = \prod_{j=1}^{n} \left[ \sum_k p_{kj} f_k(y_j) \right] \]
\[ p_{kj} = \text{Prob}\{x_j^* = k|\text{markers}\} \text{ for putative QTL} \]
\[ f_k(y_j) = \phi(k a^* + x_j^T B, V) \] multivariate normal

Maximum likelihood estimates

Expectation/Conditional Maximization (ECM) special version of general EM algorithms
Meng Rubin 1993

Expectation E-step

individual posterior QTL genotype probabilities

\[ p_{kj}^{(t+1)} = \frac{p_{kj} f_k^{(t)}(y_j)}{\sum_l p_{kj} f_l^{(t)}(y_j)} \]
\[ f_k^{(t)}(y_j) = \text{normal density functions with parameters replaced by estimates in iteration } t \]

Conditional Maximization CM-step

model parameters divided into three groups: QTL \((a^*, d^*)\), cofactors \((B)\), covariance \((V)\) estimated consecutively between groups but simultaneously within each group

log-likelihood with parameter estimates

\[ \ln(L_1(\lambda)) = -\frac{n m \ln(2\pi)}{2} - \frac{n}{2} \ln(\lambda) + \sum_{j=1}^{n} \ln \left( \sum_k p_{kj} \exp \left[ -\frac{1}{2} (y_j - k^T B) V^{-1} (y_j - k^T B) \right] \right) \]

\[ \ln(L_0(\lambda)) = -\frac{n m \ln(2\pi)}{2} - \frac{n}{2} \ln(\lambda) - \frac{n}{2} \sum_{j=1}^{n} (y_j - x_j^T B)^T V^{-1} (y_j - x_j^T B) \]

\[ |V| = \text{determinant of covariance matrix} \]

log-likelihood under null model of no QTL

QTL dropped, but cofactors remain note that covariance matrix estimate changes

\[ \ln(L_0) = -\frac{n m \ln(2\pi)}{2} - \frac{n}{2} \ln(\lambda) + \frac{n m}{2} \]
\[ \tilde{V}_0 = (Y - XB_0)^T(Y - XB_0)/n \]
\[ \tilde{B}_0 = (X^T X)^{-1} X^T Y \]
Hypothesis Tests of QTL Effects

model 1 = full model (QTL for all m traits)
model 0 = null model (no QTL)
intermediate models: QTL for only some traits
additive and/or dominance

case of m = 2 traits has key features

joint mapping for QTL on two traits

map QTL for each trait individually or jointly on both?

joint mapping hypotheses

\[ H_0 : a_1^* = 0, \ d_1^* = 0, \ a_2^* = 0, \ d_2^* = 0 \]

\[ H_1 : \text{At least one of them is not zero} \]

likelihood ratio test statistic

\[ LR_1 = -2 \ln \left( \frac{L_0}{L_1(\lambda)} \right) \]

approximately chi-square distributed under \( H_0 \)

threshold value for significance

hard to determine critical value for whole genome
same problem as CIM (Zeng 1994)
Bonferroni approximate test
extend permutation test of Churchill Doerge
available in theory–not implemented

why perform joint mapping?

formal procedures to test biologically interesting hypotheses

– pleiotropic effects of QTL
– QTL by environment interaction
– pleiotropy vs. close linkage

may perform better than separate CIM
– putative QTL has pleiotropic effects on both traits
– genotypic & environmental correlation opposite

Testing pleiotropic effects

does one QTL affect more than one trait?
pick a genome position \( \lambda \) and jointly test traits
pleiotropic effects on both traits

\[ H_{10} : a_1^* = 0, \ d_1^* = 0, \ a_2^* \neq 0, \ d_2^* \neq 0 \text{ at } \lambda \]

only trait 2 is affected

\[ H_{11} : a_1^* \neq 0, \ d_1^* \neq 0, \ a_2^* \neq 0, \ d_2^* \neq 0 \text{ at } \lambda \]

both traits affected by QTL

and

\[ H_{20} : b_1^* \neq 0, \ d_1^* \neq 0, \ b_2^* = 0, \ d_2^* = 0 \text{ at } \lambda \]

only trait 2 is affected

\[ H_{21} : b_1^* \neq 0, \ d_1^* \neq 0, \ b_2^* \neq 0, \ d_2^* \neq 0 \text{ at } \lambda \]

both traits affected by QTL

\[ H_{11} = H_{21} \] is alternative of pleiotropy
need to test \( H_{10} \text{ and } H_{20} \) together

estimates and tests under restrictions

test of \( H_{10} \text{ vs. } H_{11} \) differs from test of trait 1 alone
since traits are correlated

test has more power than separate analyses

estimates of model parameters under \( H_{10} \text{ and } H_{20} \)
use ECM with some parameters set to 0

likelihood ratio test statistics use these estimates
testing pleiotropic effects against close linkage

rejecting both $H_{10}$ and $H_{20}$ supports hypothesis of pleiotropic effects of a single QTL

what if there were two closely linked QTL?
want to separate genetic correlation from linkage

two closely linked QTL may behave like one pleiotropic QTL
one pleiotropic QTL may be estimated as two QTL with separate trait analysis

implications for genetics and breeding
power to detect the difference?
linkage vs. fine mapping: what is a QTL?

need to focus on small region for test of 2 QTL
only genome regions significant under joint mapping linkage at distance may be obvious computation costs

©ZB Zeng & BS Yandell 11.13 March 19, 2001

likelihood analysis: pleiotropy vs. close linkage

two QTL at positions $\lambda_1$, $\lambda_2$
$|\lambda_1 - \lambda_2| < 5cM$ for convenience

$H_0 : \lambda_1 = \lambda_2$
$H_1 : \lambda_1 \neq \lambda_2$

allow both QTL to have effects ($a_{ik}^1 \neq 0$)
$H_1$ is special case of many possible alternatives
more general alternative: both QTL have pleiotropic effects (more complicated)

statistical model for closely linked QTL

$$y_{j1} = b_{01} + a_{1}^1 x_{1j}^1 + \sum_l b_{l1} x_{jl} + e_{j1}$$
$$y_{j2} = b_{02} + a_{2}^2 x_{2j}^2 + \sum_l b_{l2} x_{jl} + e_{j2}$$

looks like multiple trait model defined earlier
but QTL genotypes $x_{kj}^i = x_{kj}^i(\lambda_k)$
defined for separate QTL at different positions

©ZB Zeng & BS Yandell 11.14 March 19, 2001

cautions on choice of cofactors

avoid using markers inside search region
models under hypotheses depend on cofactors

mixture model over two loci

nine components: recombination for two loci in $F_2$

$p_{kij} = \text{Prob}\{x_{1j}^1 = k, x_{2j}^2 = i|\lambda_1, \lambda_2\}$
probability $p_{kij}$ inferred from flanking markers

– different marker intervals: independence $p_{kij} = p_k p_i$
– same marker interval: Table 11.1 for 4 positions

linked QTL likelihood function

$$L_2(\lambda_1, \lambda_2) = \prod_{j=1}^n \sum_{k,i} p_{kij} f_{ki}(y_j)$$
bivariate normal density $f_{ki}(y_j)$:

$$E\left( \begin{array}{c} y_{j1} \\ y_{j2} \end{array} \right) = \left( \begin{array}{c} ka_1^1 + x_{1j}^1 b_1 \\ ia_2^2 + x_{2j}^2 b_2 \end{array} \right), \quad \text{Var}(y_j) = \mathbf{V}$$

©ZB Zeng & BS Yandell 11.15 March 19, 2001

ECM iteration to maximize likelihood

E-step: posterior probabilities of QTL genotypes

$$p_{kij}^{(t+1)} = \frac{p_{kij} f_{ki}^{(t)}(y_j)}{\sum_{k,i} p_{kij} f_{ki}^{(t)}(y_j)}$$

CM-step: maximize likelihood estimates

QTL effects

$$a_{k}^{(t+1)} = \text{blah}$$
cofactors

$$B^{(t+1)} = (X^T X)^{-1}X^T W^{(t+1)}$$

variance

$$V^{(t+1)} = \frac{(W - XB)^T (W - XB)}{n}$$

where $B = (B_1 B_2)$, $W = (W_1 W_2)$,

$$W_1 = Y_1 - (\sum_k P_k) a_1^1$$
$$W_2 = Y_2 - (\sum_i P_i) a_2^2$$

with $Y_l = \{y_{jl}\}$ and $P_k = \sum_i P_{ki}$, $P_i = \sum_k P_{ki}$

©ZB Zeng & BS Yandell 11.16 March 19, 2001
joint log likelihood for linked QTL

\[
\ln(L_2(\lambda_1, \lambda_2)) = \frac{-nm \ln(2\pi)}{2} - \frac{n}{2} \ln(V) + \sum_{j=1}^{n} \left\{ \sum_{k} p_{kj} \exp \left[ \frac{1}{2}(y_j - \hat{u}_{kj})V^{-1}(y_j - \hat{u}_{kj})^T \right] \right\}.
\]

with \( \hat{u}_{kj} = E(y_j) \)

search possible \( \lambda_1, \lambda_2 \) in region

test statistic

\[
LR_2 = -2 \ln \left( \frac{\max_{\lambda_1, \lambda_2} L_2(\lambda_1, \lambda_2)}{\max_{\lambda_1, \lambda_2} L_2(\lambda_1, \lambda_2)} \right)
\]

nested hypotheses: asymptotically \( \chi^2 \) under \( H_0 \)

scan LODs for joint and separate QTL

approximate test: do peaks match?

grid search in neighborhood

QTL by environment interaction

different environments \( \rightarrow \) different gene effects

Paterson et al. (1991); Stuber et al. (1992)

Design I: same genotypes evaluated in different environments (paired comparison)

Design II: different genotypes (individuals) from common population evaluated in different environments (group comparison)

QTL \( \times \) environment interaction hypotheses

\[
H_0: a_1^* = a_2^* = a^*, \quad d_1^* = d_2^* = d^*
\]

\[
H_1: a_1^* \neq a_2^*, \quad d_1^* \neq d_2^*
\]

only test in regions suggested by joint mapping (why?)

recombination probabilities

\[
p_{ki,j} = \text{Prob}(x_{i,j}^* = i), \quad k = 1, 2, \ j = 1, \ldots, n_k
\]

Design I paired comparison

same X (marker data) matrix

multiple phenotypic vectors \( Y \) across environments

same statistical model as multiple traits

\( V \) reflects within and between environment variation

log likelihood under \( H_0 \): not \( G \times E \)

construct likelihood \( L_3 \) under restriction of \( H_0 \)

maximize likelihood using ECM again

E-step: substitute \( a^* \) for \( a_1^*, a_2^* \) (\( d^* \) for \( d_1^*, d_2^* \))

CM-step: \( a^{(t+1)} \) is \( V^{(t)} \) weighted average

\[
V = \begin{pmatrix} \sigma_1^2 & \rho \sigma_1 \sigma_2 \\ \rho \sigma_1 \sigma_2 & \sigma_2^2 \end{pmatrix}
\]

likelihood ratio test for \( G \times E \)

\[
LR_3(\lambda) = -2 \ln(L_3(\lambda)/L_1(\lambda))
\]

asymptotically chi-square under \( H_0 \)

Design II group comparison

statistical model

\[
y_{i,j} = x_{i,j}^1 a_1^* + x_{i,j}^1 b_1 + e_{i,j} \quad j = 1, 2, \ldots, n_1
\]

\[
y_{i,j} = x_{i,j}^2 a_2^* + x_{i,j}^2 b_2 + e_{i,j} \quad j = 1, 2, \ldots, n_2
\]

matrix notation

\[
y_1 = x_1^1 a_1^* + X_1 b_1 + e_1
\]

\[
y_2 = x_2^2 a_2^* + X_2 b_2 + e_2
\]

assume environmental effects \( e_{i,j} \) independent normal with means zero and variances \( \sigma_1^2, \sigma_2^2 \)

estimate separately by environment under \( H_1 \)

sum of separate \( \ln(L_1) \)s by environment

\[
\ln(L_4(\lambda)) = \sum_{j=1}^{n_1} \ln \left( \sum_{k} p_{kj} f(y_{i,j}) \right) + \sum_{j=1}^{n_2} \ln \left( \sum_{k} p_{kj} f(y_{i,j}) \right)
\]

\[
L_{11}(\lambda), L_{12}(\lambda) \text{ are } L_1(\lambda) \text{ for groups 1,2}
\]
Design II group comparison
estimate jointly under $H_0$

one QTL effect parameter $a^*$
same $\lambda$, different individuals $1_j$ and $2_j$
$p_{1ij}$ and $p_{2ij}$ independent but posterior probabilities
depend on $a^*$ in E-step through normal density

$$P_{kij}^{(t+1)} = \frac{p_{kij}f_i^{(t)}(y_{kj})}{\sum_{i=0}^2 p_{kij}f_i^{(t)}(y_{kj})}, \ k = 1, 2$$

CM-step involves block update
– QTL effect $a^*$
– cofactors $B_1, B_2$
– variances $\sigma_1^2, \sigma_2^2$

$\ln(L_5(\lambda))$ looks like $\ln(L_4(\lambda))$ with $\tilde{a}_1^* = \tilde{a}_2^* = \tilde{a}^*$
likelihood ratio test statistic

$$LR_4(\lambda) = -2 \ln(L_5(\lambda)/L_4(\lambda))$$
asymptotically chi-square under $H_0$: no $G \times E$
degrees of freedom depend on model (BC, $F_2$)

relative efficiency of Designs I and II

mapping QTL, testing QTL $\times$ environment interaction
assume $n_1 = n_2 = n$ and $n$ large

$LR_4(\lambda)$ is special case of $LR_3(\lambda)$ with $\rho = 0$
Design II: has more power for mapping QTL
Design I: more power to detect QTL $\times$ env interaction

QTL $\times$ environment as fixed effects here
random effects $\rightarrow$ mixed models

@ZB Zeng & BS Yandell 11.21 March 19, 2001