- Project presentations and papers due:
  Thursday, MARCH 15, 2007,
  1030-1220

In Genome Sciences Foege S-110

Lab section today in Genome Sciences Foege S-110

Reading

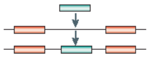
- Today
  - Jingwei Nature paper
  - Nature reviews genetics review of new genes

Origin of new genes

- Overview of processes generating new genes
- Example of jingwei
  - HKA test
  - MK test

Why “new” genes?

- Do not want events obscured by evolutionary divergence
- Gives ability to reconstruct history.

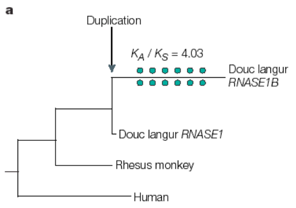
Mechanism	Process	Examples
Exon shuffling: ectopic recombination of exons and domains from distinct genes		*fucosyltransferase*, *jingwei*, *Tre2*

Exon shuffling problem common in gene evolution.
Easier to modify exon then generate one de novo.
Will discuss jingwei in detail.

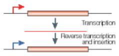
Gene duplication: classic model of duplication with divergence		*CG*β, *Cid* , *RNASE1B*

We discussed lysozyme, DDC model

Rnase1 also involved in foregut fermentation



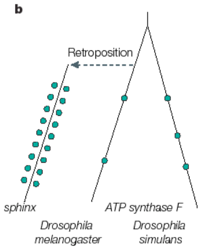
Retroposition:
new gene duplicates
are created in new
genomic positions
by reverse transcription
or other processes



PGAM3, *Pgk2*, *PMCHL1*,
PMCHL2, *Sphinx*

1% of human DNA is
retroposed to new
genomic locations

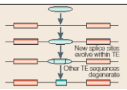
8



What would be a signal of a retroposed gene?

How are retroposed genes regulated?

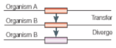
Mobile element:
a mobile element,
also known as a
transposable element
(TE), sequence is
directly recruited by
host genes



HLA-DR-1, human *DAF*,
lungerkine mRNA,
mNSC1 mRNA

Generates 4% of new
exons in human
protein-coding genes

Lateral gene transfer:
a gene is laterally
(horizontally) transmitted
among organisms



acytylneuraminate lysase,
Escherichia coli
mutU and *mutS*

Most often reported
in prokaryotes and
recently reported
in plants

Gene fusion/fission:
two adjacent genes
fuse into a single gene,
or a single gene splits into
two genes

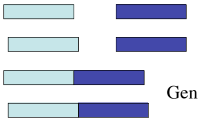
Fatty-acid synthesis enzymes,
Kua-UEV, *Sdic*

Involved in the
formation of ~0.5%
of prokaryotic genes

Gene fusion / fission used to predict protein-protein interactions

Example, two enzymes form dimer
to catalyze reaction

Focus species



Species 1 — Genes separate

Species 2

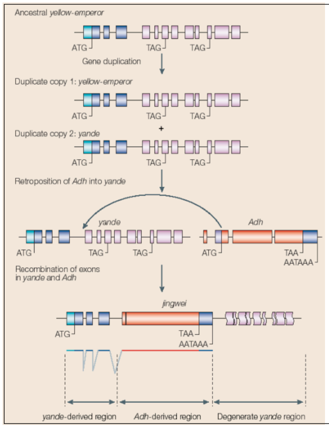
Species 3 — Genes fused

Species 4

De novo origination:
a coding region
originates from a
previously non-coding
genomic region

Any examples?

Evolution of new gene jingwei



Is Jingwei a pseudogene? What are expectations?

- Polymorphisms
- Rates of evolution
- Stop codons
- Insertions / deletions

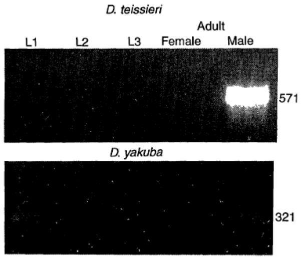
Is Jinwei a pseudogene?

- Polymorphisms
  - Most silent
- Rates of evolution
  - Lots of NS between jgw and ADH
- Stop codons
  - No new stop codons
- Insertions / deletions
  - No length polymorphisms in coding region

Is Jingwei expressed?

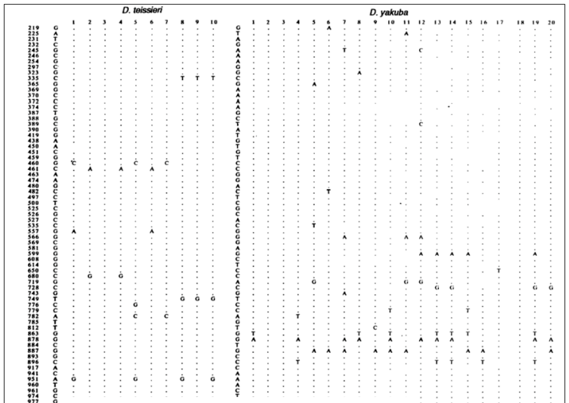
- Northern blots showed only ADH
  - Jgw may be expressed at low levels
- RT-PCR performed and found it to be expressed.

Adult male expression.
This is same expression pattern as yande and yellow-emperor.



Population level to test for adaptive evolution

- Compare rates of evolution within and between species
  - Under neutrality, these rates should be correlated.



What would be a signal of selective sweep?

S (segregating sites)

- The total number of variable sites in sample
- Sometimes divided by number of sites to give segregating sites per nucleotide.

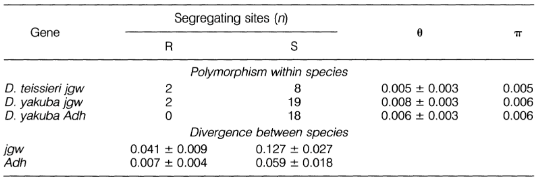
Pi (π)

- Average number of pairwise differences divided by length analyzed.

$$\Pi = \frac{1}{[n(n-1)/2]} \sum_{i<j} \Pi_{ij}$$

Theta (θ)

- Number polymorphic nucleotides, normalized for sample size
- Should be similar to π
  - Remember Joanna's talk



Gene	Segregating sites (*n*)		θ	π
	R	S		
	Polymorphism within species			
D. teissieri jgw	2	8	0.005 ± 0.003	0.005
D. yakuba jgw	2	19	0.008 ± 0.003	0.006
D. yakuba Adh	0	18	0.006 ± 0.003	0.006
	Divergence between species			
jgw	0.041 ± 0.009	0.127 ± 0.027		
Adh	0.007 ± 0.004	0.059 ± 0.018		

What tests could they have done with polymorphism data?

Look at jingwei vs Adh

- 8 fixed differences in Adh region
- All nonsynonymous

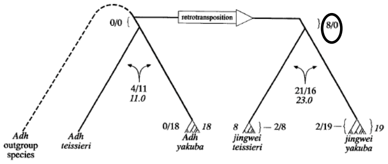


Fig. 4. Phylogenetic analysis of the evolution of the *Adh*-derived portion of *jingwei*. The topology reflects the assumed evolutionary relation between *jingwei* and *Adh* (*5*). The small (stippled) trees at the tips indicate that the detailed phylogenies of the alleles within species are unknown. Numbers presented as fractions are the numbers of (fixed or polymorphic) amino acid replacement differences over the numbers of (fixed or polymorphic) silent differences. The *Adh* outgroup includes the sequences of *D. melanogaster, D. simulans, D. sechellia, D. mauritiana, D. erecta*, and *D. orena*. These outgroup sequences were used to confirm that all functional *Adh* genes have the same DNA sequence for the eight codons that are distinct to *jingwei* (in both *D. teissieri* and *D. yakuba*) with one exception. At the eighth codon (nucleotide position 954) of *D. orena Adh*, there is a substitution that conferred an amino acid replacement. The italic numbers are the average numbers of silent substitutions and the numbers of silent polymorphisms under the HKA test, which also incorporates the standard assumptions of the neutral theory.

Is there some way to perform a statistical test of this divergence?

Now we have lots of polymorphism data. How do we test for selection?

HKA test (handout)
Hudson, Kreitman Aquade

- **Within species variation and between species variation in different genes or gene regions should be similar**
- Assumes neutral mutation rate constant
- Assumes population sizes have not changed
- For Jingwei, focus on syn. Changes since already showed elevated nonsyn.
  - Hitchhiking

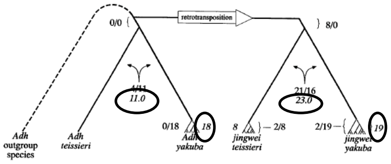


Fig. 4. Phylogenetic analysis of the evolution of the *Adh*-derived portion of *jingwei*. The topology reflects the assumed evolutionary relation between *jingwei* and *Adh* (*5*). The small (stippled) trees at the tips indicate that the detailed phylogenies of the alleles within species are unknown. Numbers presented as fractions are the numbers of (fixed or polymorphic) amino acid replacement differences over the numbers of (fixed or polymorphic) silent differences. The *Adh* outgroup includes the sequences of *D. melanogaster, D. simulans, D. sechellia, D. mauritiana, D. erecta*, and *D. orena*. These outgroup sequences were used to confirm that all functional *Adh* genes have the same DNA sequence for the eight codons that are distinct to *jingwei* (in both *D. teissieri* and *D. yakuba*) with one exception. At the eighth codon (nucleotide position 954) of *D. orena Adh*, there is a substitution that conferred an amino acid replacement. The italic numbers are the average numbers of silent substitutions and the numbers of silent polymorphisms under the HKA test, which also incorporates the standard assumptions of the neutral theory.

	ADH	*Jgw*
Polymorphic	18	19
Divergent	11	23

Similar levels of polymorphism
But divergence 2X higher in jgw

Only used *D. yakuba*

MK test
MacDonald - Kreitman

- **Ratio of nonsynonymous to synonymous changes should be similar within and between species**
- Assumes neutral mutation rate constant
- Does not require population levels to be constant.
  - Robust to demographic effects

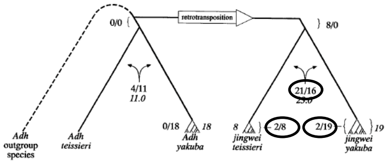


Fig. 4. Phylogenetic analysis of the evolution of the *Adh*-derived portion of *jingwei*. The topology reflects the assumed evolutionary relation between *jingwei* and *Adh* (*5*). The small (stippled) trees at the tips indicate that the detailed phylogenies of the alleles within species are unknown. Numbers presented as fractions are the numbers of (fixed or polymorphic) amino acid replacement differences over the numbers of (fixed or polymorphic) silent differences. The *Adh* outgroup includes the sequences of *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. erecta*, and *D. orena*. These outgroup sequences were used to confirm that all functional *Adh* genes have the same DNA sequence for the eight codons that are distinct to *jingwei* (in both *D. teissieri* and *D. yakuba*) with one exception. At the eighth codon (nucleotide position 954) of *D. orena Adh*, there is a substitution that conferred an amino acid replacement. The italic numbers are the average numbers of silent substitutions and the numbers of silent polymorphisms under the HKA test, which also incorporates the standard assumptions of the neutral theory.

MK Test: Basically asking is ratio NS/S is same within (poly) and between species (fixed)

	Fixed	sum Poly	tes Poly	yak Poly
NonSyn	21	4	2	2
Syn	16	27	8	19

Issue of determining if excess between species or within.

Comparison to tes p = .05

Comparison to yak < .001

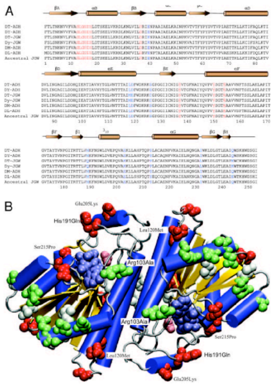
Summary of jengwei

- Previous functions remained intact
- Chimeric structures can generate new diversity
- Remainder of fusion degenerated
- Test of selection
  - 8 NS changes
    - No test of significance
  - HKA test
    - Only tested yakuba
  - MK test
    - Mainly driven by yakuba, but teissieri still significant

Why difference? Is there selection on teis? Expression differences?

Is this adaptive evolution associated with functional divergence

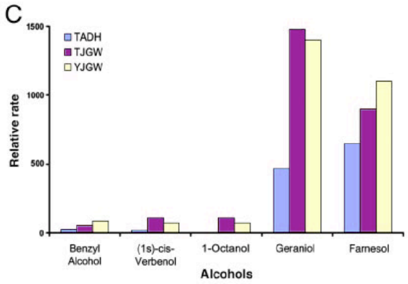
- Map onto 3D structure
  - ADH structure known, model jgw
- Measure specificity
  - Express jgw and ADH, measure activity on different substrates



Only 2 of 8 fall near catalytic domain - but could others change specificity?

Check for functional divergence

- Express jingwei, Adh
- Measure alcohol specificity.



Jingwei evolved to utilize long-chain primary alcohols used in hormone and pheromone metabolism.

Conclusion

- Jingwei is a new gene generated by a variety of processes
- Adaptive evolution following creation of gene - what species?
- Adaptive evolution resulted in switch of alcohol specificity.

questions

- Describe 3 ways new genes can evolve
- What is the basis of the MK test?
- What is the basis of the HKA test?
- What type of mutations would you expect to observe in pseudogenes?
- What are singletons, and how can they be used to detect selection?