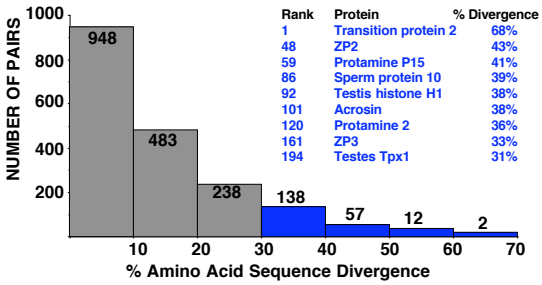
Speciation by the rapid evolution of reproductive proteins

Willie J. Swanson

University of Washington
Department of Genome Sciences

Rodent x Human 1880 Orthologous Sequence Pairs (~4% of genes)

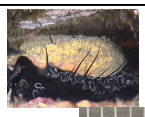


Rank	Protein	% Divergence
1	Transition protein 2	68%
48	ZP2	43%
59	Protamine P15	41%
86	Sperm protein 10	39%
92	Testis histone H1	38%
101	Acrosin	38%
120	Protamine 2	36%
161	ZP3	33%
194	Testes Tpx1	31%

Makalowski & Bogusti, *PNAS* 95, 9407 (1998)

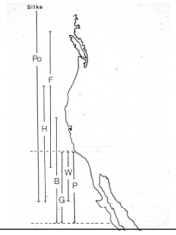
Three groups of organisms studied

- **Marine invertebrates**
  - **Abalone**
- **Terrestrial vertebrates**
  - **Mammals**
- ***Drosophila***
  - **Fruit flies**

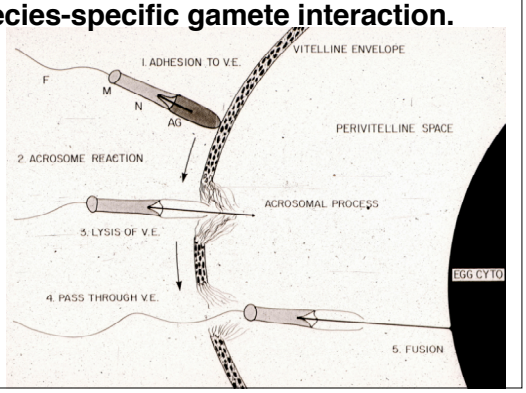


Abalone (Vacquier Lab)

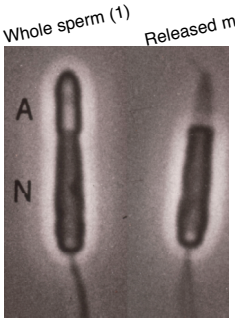
- **Gastropod mollusk, genus *Haliotis.***
- **~70 species worldwide, 7 off Eastern Pacific.**
- **Free spawning, external fertilization.**
- **Hybrids rare.**

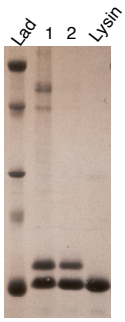


Species-specific gamete interaction.

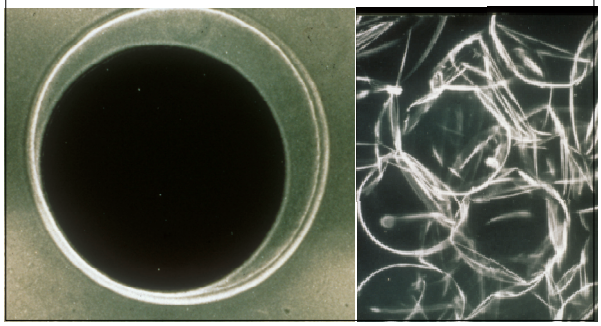


Isolation of sperm acrosomal proteins

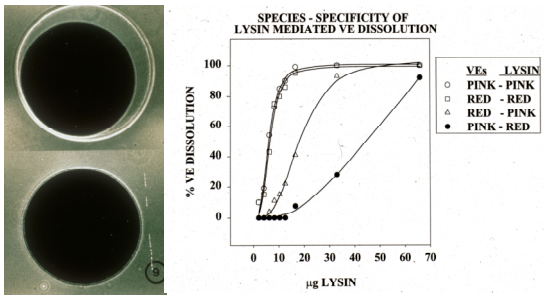




Isolation of egg vitelline envelopes (VEs)

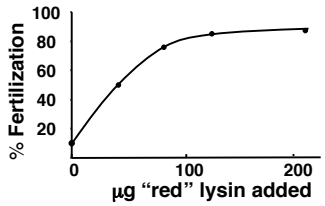


The sperm acrosomal protein lysin species-specifically dissolves the egg vitelline envelope (VE)



Dissolution of VE by lysin is a species-specific barrier to fertilization

Cross between "red" egg and "green" sperm



Increased hybrid fertilization by removing VE with homospecific lysin

Lysin

- 27 species sequenced.
- 3 crystal structures.
- Extensive biochemistry.
- Evolutionary analyses.
  - Rapid evolution and positive selection.

→ Vacquier Lab SIO/UCSD

Exons evolve up to 16X faster than introns

Table 2 | **Percentage sequence difference* in three abalone species**

Gene	Percentage nucleotide difference		
	Hru–Hco	Hru–Hfu	Hco–Hfu
Lysin			
Exons (420 bp)	13.7	24.1	22.1
Introns (2187 bp)	3.0	4.8	5.8
sp18			
Exons (447 bp)	83.6	81.3	92.8
Introns (745 bp)	ND	5.1	ND

Ed Metz & Cuquis Robles
PNAS 95: 10676, 1998

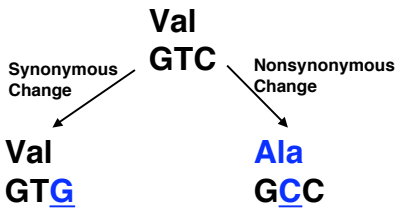
Why is lysin different between species?

- Lack of constraint:
  Selectively neutral evolution
- Adaptive value for change:
  Positive "Darwinian" selection

→ Compare cDNA sequences.

2 types of changes in codons

Synonymous = silent change (amino acid stays the same)
Nonsynonymous = replacement change (changes amino acid)

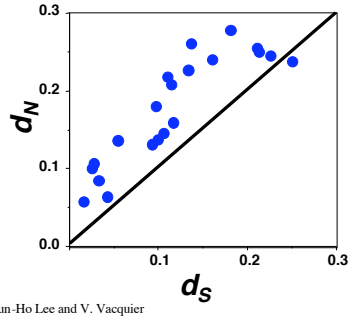


d_N = # nonsynonymous substitutions/# nonsynonymous sites
d_S = # synonymous substitutions/# synonymous sites

Test for selection by comparing d_N and d_S
$d_N/d_S = 1$: Neutral evolution
$d_N/d_S < 1$: Purifying selection
$d_N/d_S > 1$: Positive selection

The d_N/d_S ratio (ω) measures the selective pressure

Abalone sperm lysin shows signs of selection in pairwise comparisons between 7 species



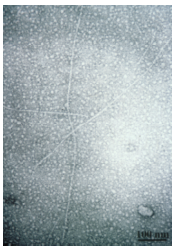
Data from Youn-Ho Lee and V. Vacquier

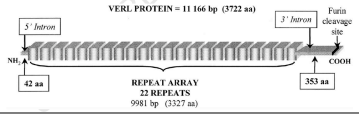
Why is lysin rapidly evolving?

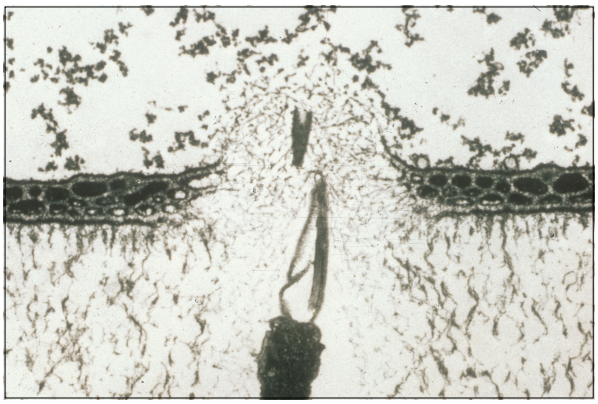
- Identify the interacting VE component.
- Study evolution of VE receptor for lysin (VERL).
  - Is the female component also subjected to positive selection?

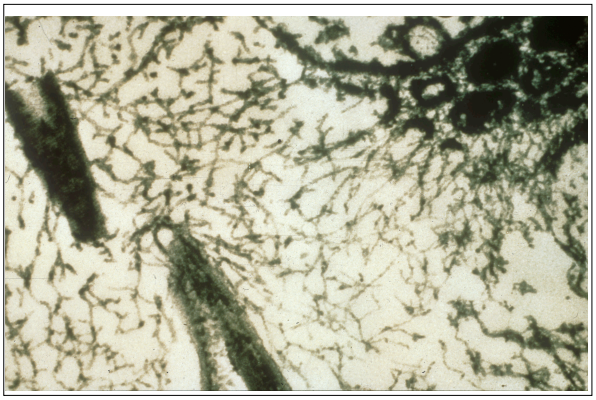
Characterization of VERL

- 1,500 kDa.
- Species-specific binding to lysin
  (EC_{50} ~9.5 nM, Hill Coefficient ~3).
- 22 repetitive domains.
- Repeats 3-22 are 95% identical.
  (homogenized by concerted evolution)







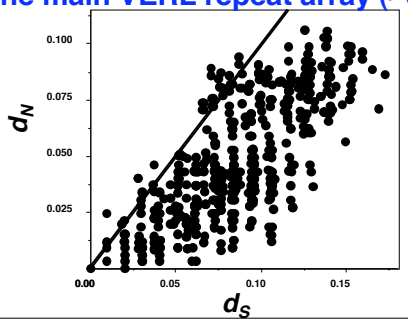


Is egg VERL subjected to

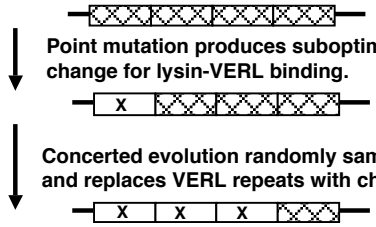
positive selection like sperm lysin?

- Estimate d_N/d_S ratios

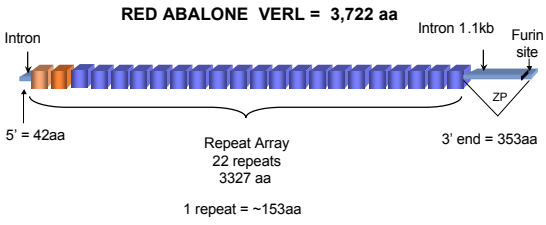
No sign of positive selection for repeats in the main VERL repeat array (~3-22)



Repetitive nature of VERL may lead to neutral evolution of each individual repeat unit

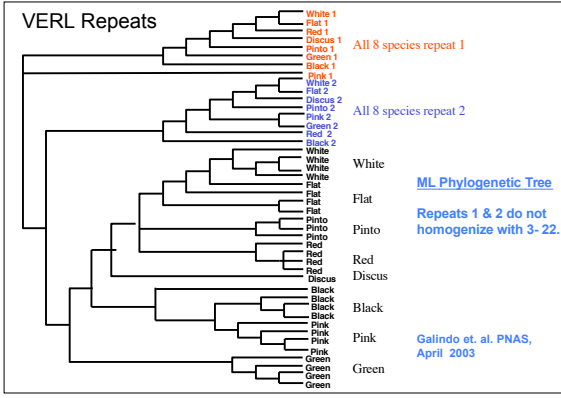


Positive selection occurs on lysin to maintain interaction.

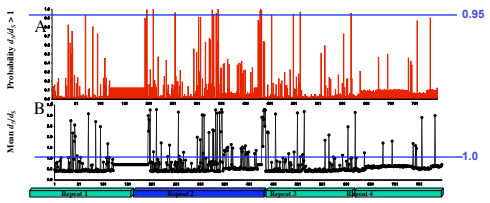


Repeats 3-22 are >95% identical to each other and are under relaxed selection with $d_N/d_S \sim 1$.

Repeats 1 and 2 evolve independently and are not homogenized and are subjected to positive selection with $d_N/d_S > 1$.

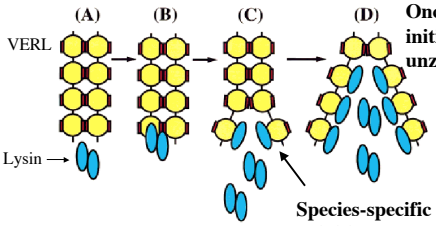


ML Analysis of VERL Repeats 1- 4 From Eight Species. Positive Selection is Most Robust in Repeat 2



Galindo, et.al., PNAS April, 2003

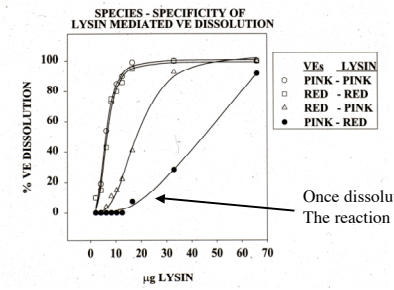
Model for species - specific dissolution of VE by lysin



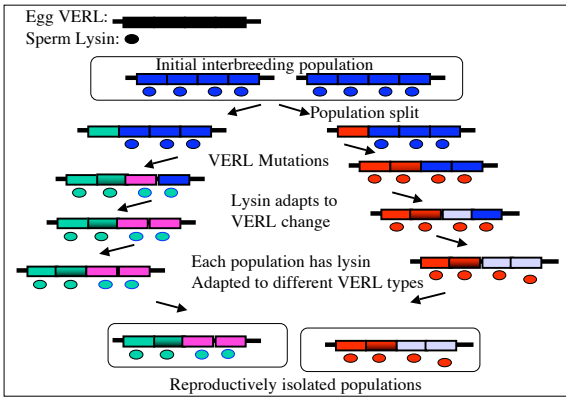
Once unraveling initiated, rapid unzipping of fibers.

Species-specific step to initiate unraveling of VERL fibers.

Model of repeats 1 & 2 regulating species - specificity is consistent with VE dissolution kinetics.

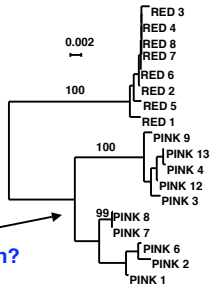


Once dissolution initiated, The reaction proceed rapidly



If rapid evolution of fertilization proteins is involved in speciation, we should observe differentiation WITHIN a species at these loci.

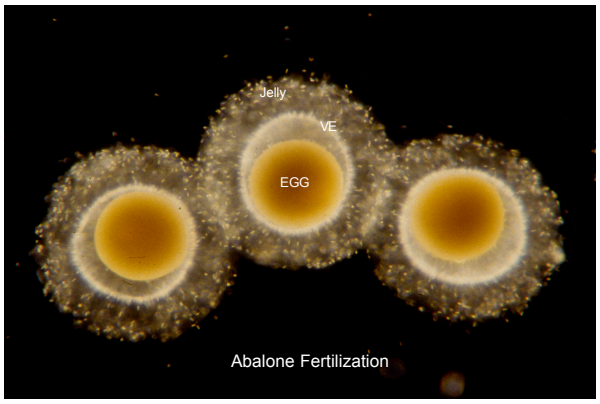
Abalone VERL polymorphism



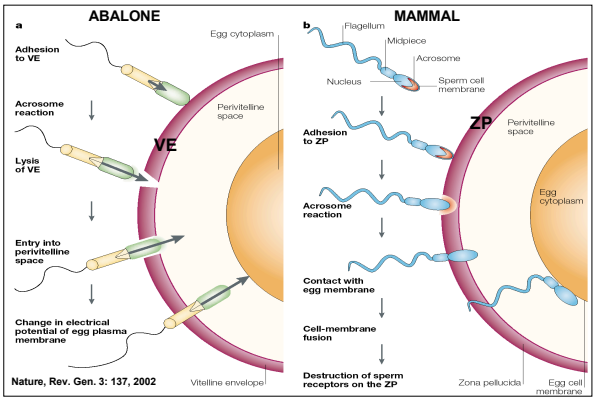
Recent (incipient) speciation?

Hypotheses for rapid divergence

- **Sexual conflict (polyspermy, gamete usage)**
- **Sperm competition**
- **Gene duplication/functional diversification**
- **Relaxed constraint/repetitive domains**
- **Selection against hybrids (reinforcement)**
- **Sexual selection**
- **Cryptic female choice**
- **Self-nonself recognition**
- **Barrier to microbes/immunological defense**



Is adaptive evolution of reproductive proteins limited to free-spawning marine invertebrates?



Egg Coat Zona Pellucida (ZP) Glycoproteins

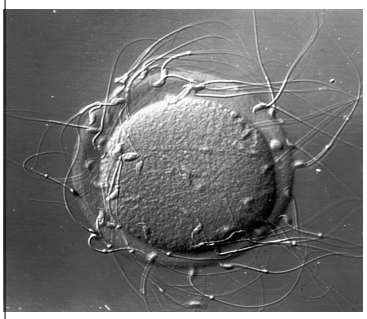
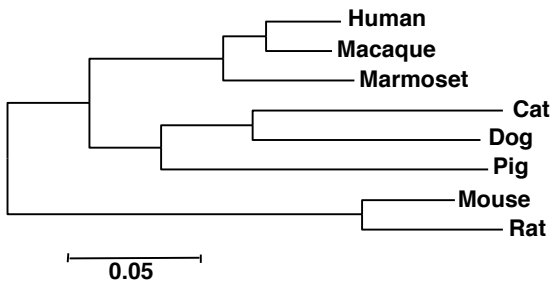


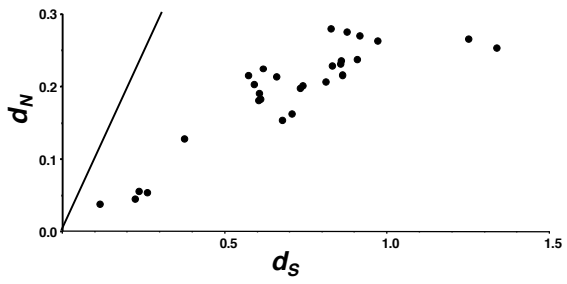
Photo from Wassarman, 1999

- ZP2 and ZP3 are among the 10% most rapidly evolving proteins between humans and rodents

Phylogenetic tree of *ZP3* genes used in analysis



ZP3 d_N/d_S ratio averaged across all sites



d_N/d_S ratio (ω) estimated across all sites is insensitive at detecting positive selection

e.g. 3/44 (6.8%) sites show signs of positive selection

```
1 MSLAVLTFLVLCGFSFQHQAVGKWLTAAQKHPISGRMIRIRTKE
2 MSLAVLTFLVLCGWSFQHQAVGKWLTLAQKHPISGKMIRIRTKE
3 MSLAVLTFLVLCGYSFQHQAVGKWLTSAQKHPISGHMIRIRTKE
```

Variation between sites analyzed by likelihood ratio tests.

Compare likelihood of neutral vs. selection models[1]

1st: Identification of positive selection

- Is there a class of sites with $d_N/d_S > 1$?

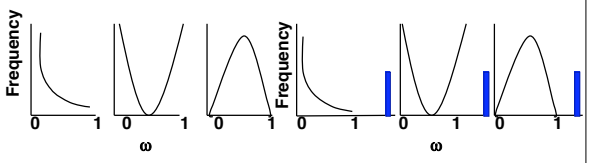
2nd: Identification of sites with $d_N/d_S > 1$.

- Where are the sites subjected to positive selection?

[1]Method of Nielsen and Yang (1998), Yang et al. (2000)

Variation between sites analyzed by likelihood ratio tests. Compare likelihood of neutral vs. selection models[1].

Neutral model	Selection model
Assume beta distribution of ω in interval (0,1)	Beta distribution, ω_S estimated



[1]Method of Nielsen and Yang (1998), Yang et al. (2000)

1st: Testing for positive selection using likelihood ratio test statistic:

$$\Delta l = \log\left(\frac{\max\{L(\text{neutral model})\}}{\max\{L(\text{selection model})\}}\right)$$

$$= \log(\max\{L(\text{neutral model})\}) - \log(\max\{L(\text{selection model})\})$$

$-2\Delta l$ approximates χ^2 with n degrees freedom, where n is the difference in number of parameters between the nested models.

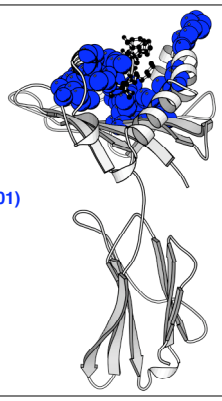
2nd: Identifying sites subjected to selection

If likelihood ratio test shows a significant difference, then positive selection is indicated. Given the distribution of ω, we then use an empirical Bayes approach to predict sites subjected to positive selection.

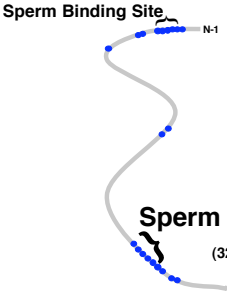
Control Genes for Tests of Positive Selection

MHC class I is known to be under positive selection
Positive selection detected
7% sites with $d_N/d_S = 4.0$ ($p<0.01$)

"Housekeeping" gene (Carbonic Anhydrase) is not under positive selection.
No positive selection detected



Mammalian Egg Coat Protein ZP3



Sites under selection in *ZP3* correspond to region identified to be involved in species-specific sperm-egg interaction
8% of sites with $d_N/d_S = 2$ ($p<0.05$)

Potential effects of rapid evolution

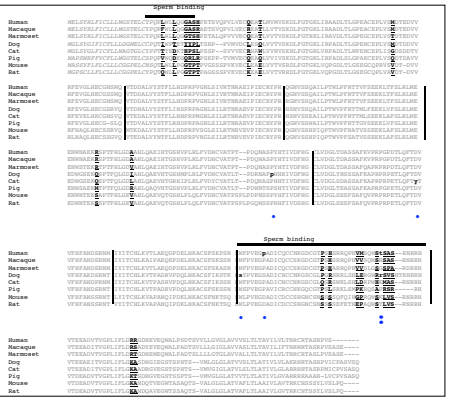
- **Rapidly evolving ZP2 and ZP3**
  - **Could a mismatch of sperm/egg proteins contribute to infertility?**
- **Rapidly evolving MHC genes**
  - **Exact match of tissue-type needed for successful skin grafts.**
- **Variable blood groups (surface antigens)**
  - **Blood types must match to serve as blood donors.**

Is there any evidence that reproductive proteins show variation within humans?

ZP3 Polymorphism in database (errors?)

-sequence human populations (47) to determine extent of AA polymorphism

(Nathan Clark)



Future plans: Clinical Association Study

(“evolutionary medicine”)

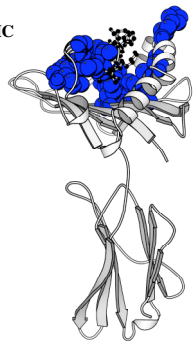
- Use clinical data to obtain samples where *in vitro* fertilization fails, but sperm and egg are normal.
- Genotype sperm and egg molecules
- Perform association study to determine if any genotypes result in decreases IF success.
  - If additional data available on efficiency of IF (ie sperm concentration), incorporate probabilities.
- Develop diagnostic test to indicate when to skip IF in favor of ICSI.

Rapid, adaptive evolution of reproductive proteins occurs in several taxonomic groups.

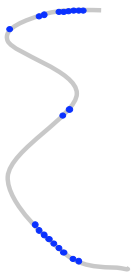


Functional insights from amino acid divergence

MHC



ZP3



The selective pressure(s) remains unknown, but could be involved in speciation.

Hypotheses for rapid divergence

- **Sexual conflict (polyspermy, gamete usage)**
- **Sperm competition**
- **Gene duplication/functional diversification**
- **Relaxed constraint/repetitive domains**
- **Selection against hybrids (reinforcement)**
- **Sexual selection**
- **Cryptic female choice**
- **Self-nonself recognition**
- **Barrier to microbes/immunological defense**