MOLECULAR AND CELLULAR BIOLOGY, Aug. 1988, p. 3094–3103
0270-7306/88/083094-10$02.00/0

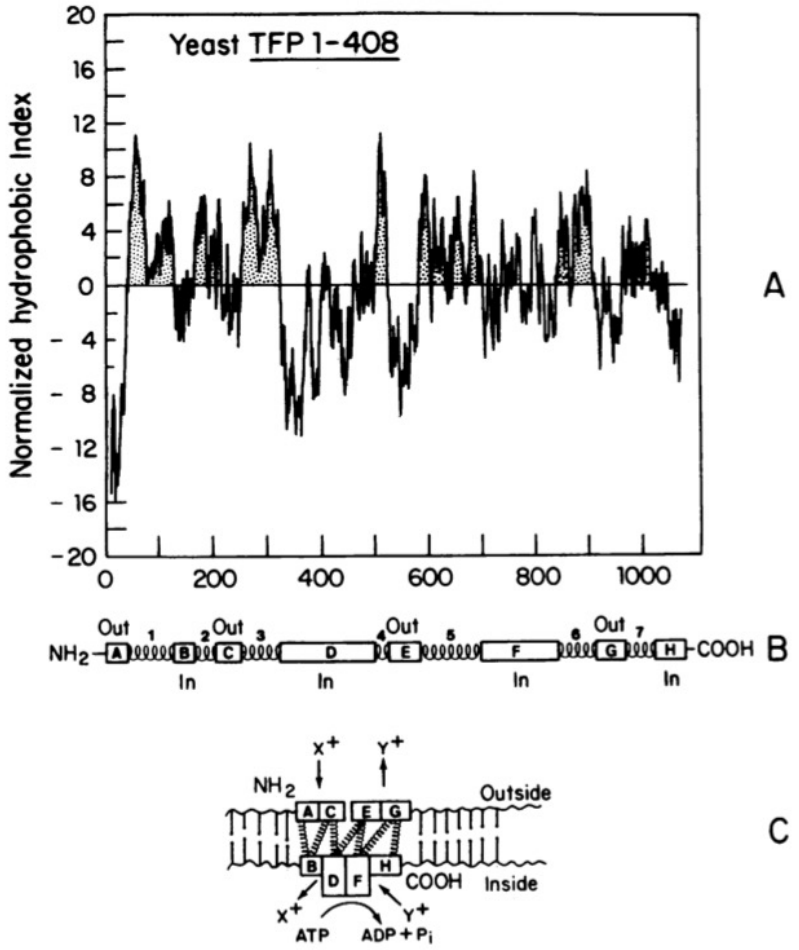
Vol. 8, No. 8

A Dominant Trifluoperazine Resistance Gene from *Saccharomyces cerevisiae* Has Homology with F_0F_1 ATP Synthase and Confers Calcium-Sensitive Growth

CHENG-KON SHIH, RONALD WAGNER, SABINE FEINSTEIN, CYNTHIA KANIK-ENNULAT, AND NORMA NEFF*

First published intein sequence, but not recognized as such.

(vacuolar ATPases catalytic subunit is not an integral membrane protein)



THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 265, No. 12, Issue of April 25, pp. 6726–6733, 1990
Printed in U.S.A.

Molecular Structure of a Gene, *VMA1*, Encoding the Catalytic Subunit of H$^+$-Translocating Adenosine Triphosphatase from Vacuolar Membranes of *Saccharomyces cerevisiae**

(Received for publication, October 10, 1989)

Ryogo Hirata, Yoshinori Ohsumi‡, Akihiko Nakano, Hiroshi Kawasaki§, Koichi Suzuki§, and Yasuhiro Anraku¶

From the Department of Biology, Faculty of Science, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan and the §Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Tokyo 113, Japan

```
S.cerevisiae (1071aa)   MAGAIENARKEIKRISLEDHAESEYGAIYSVSGPVVIAEN
                                                    * *********.**.
N.crassa     ( 607aa)                MAPQQNGAEVDGIHTGKIYSVSGPVVVAED

 41' MIGCAMYELVKVGHDNLVGEVIRIDGDKATIQVYEETAGLTVGDPVLRTGKPLSVELGPG
     *** ***********.********.**.***********. ********************
 31' MIGVAMYELVKVGHDQLVGEVIRINGDQATIQVYEETAGVMVGDPVLRTGKPLSVELGPG

101' LMETIYDGIQRPLKAIKEESQSIYIPRGIDTPALDRTIKWQFTPGKFQVGDHISGGDIYG
     *...*********. * *.*.********.******. **.*** ...*****.***..*
 91' LLNNIYDGIQRPLEKIAEASNSIYIPRGIATPALDRKKKWEFTP-TMKVGDHIAGGDVWG

161' SVFENSLISSHKILLPPRSRGTITWIAPAGEYTLDEKILEVEFDGKKSDFTLYHTWPVRV
     .*.***.** ********.*****.**  ****..************..... .******
150' TVYENSFISVHKILLPPRARGTITRIAEKGEYTVEEKILEVEFDGKKTEYPMMQTWPVRV

221' PRPVTEKLSADYPLLTGQRVLDALFPCVQGGTTCIPGAFGCGKTVISQSLSKYSNSDAII
     ***..** ** . * *.**********.*****. ***************.**.****.*.
210' PRPAAE SANQPFLT RVLDALFPSVQGGTVAIPGAFGCGKTVISQSVSKFSNSDVIV

281' YVGC ..(454a.a.)... GERGNEMAEVLMEFPELYTEMSGTKEPIMKRTTLVANTSN
     **** ***********.**********.**** .*..* **********.*****
270' YVGC ----------- -GERGNEMAEVLKDFPELSIEVDGRKEPIMKRTTLIANTSN

779' MPVAAREASIYTGITLAEYFRDQGKNVSMIADSSSRWAEALREISGRLGEMPADQGFPAY
     ***************.********.**.*.****************************
314' MPVAAREASIYTGITVAEYFRDQGMNVAMMADSSSRWAEALREISGRLGEMPADQGFPAY

839' LGAKLASFYERAGKAVALGSPDRTGSVSIVAAVSPAGGDFSDPVTTATLGITQVFWGLDK
     **************. ***** *.******.****.*********.*****.********
374' LGAKLASFYERAGKVQALGSPPREGSVSIVGAVSPPGGDFSDPVTSATLGIVQVFWGLDK

899' KLAQRKHFPSINTSVSYSKYTNVLNKFYDSNYPEFPVLRDRMKEILSNAEELEQVVQLVG
     ********************  ..*.*.*...**.** ****....**..***.*******
434' KLAQRKHFPSINTSVSYSKYLTILDKWYEREYPDFPRLRDRIRQLLSDSEELDQVVQLVG

959' KSALSDSDKITLDVATLIKEDFLQQNGYSTYDAFCPIWKTFDMMRAFISYHDEAQKAVAN
     ******.******.***************.**.*******  **. ....*******.*.
494' KSALSDPDKITLDMATLIKEDFLQQNGYSDYDQFCPIWKTEWMMKLHMGFHDEAQKAIAQ

1019' GANWSKLADSTGDVKHAVSSSKFPEPSRGEKEVHGEFEKLLSTMQERFAESTD
      *.**.*. ..* *.. ...* **  ** *.... ..* .  * ..**. .*
 554' GQNWNKVREATQDLQAQLKSLKFEVPSEGQEKICKKYEAIQQQMLDKFASVIDE
```

FIG. 6. **Amino acid sequence homology between the catalytic subunits of yeast and *Neurospora crassa* vacuolar membrane H$^+$-ATPases.** Identical residues and conserved amino acid replacements are indicated by *stars* and *dots*, respectively. *Dashes* represent gaps introduced to obtain maximum matching.

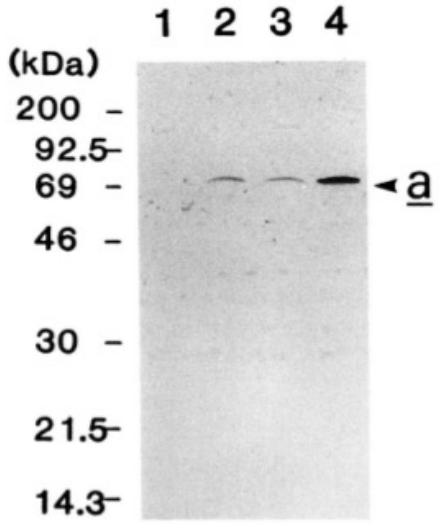


FIG. 5. **Lack of subunit *a* in total cell lysate from *vma1* null mutant cells.** Total cell lysate was prepared as described under "Materials and Methods" from RH101 (*vma1*, *lane 1*), ANY21 (*VMA1*, *lane 2*) and ANY21 harboring multicopy plasmid without or with *VMA1* gene (*lanes 3* and *4*). About 50 μg of proteins was separated on a 10% SDS-polyacrylamide gel and blotted onto a nitrocellulose membrane. Subunit *a* was detected using an anti-subunit *a* monoclonal antibody.

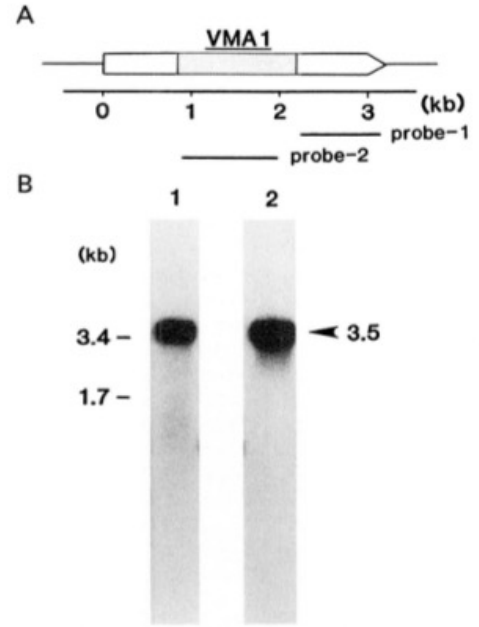
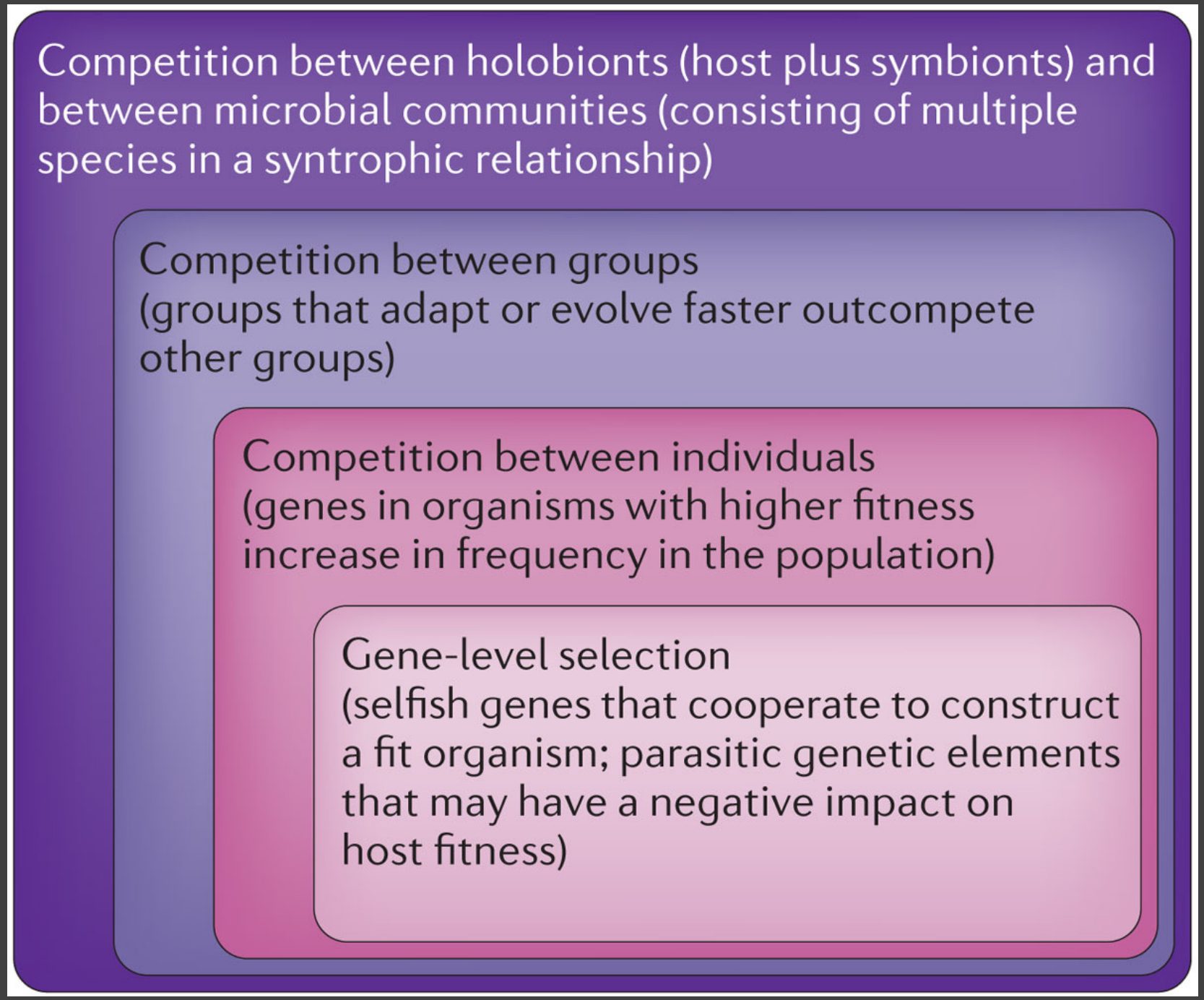


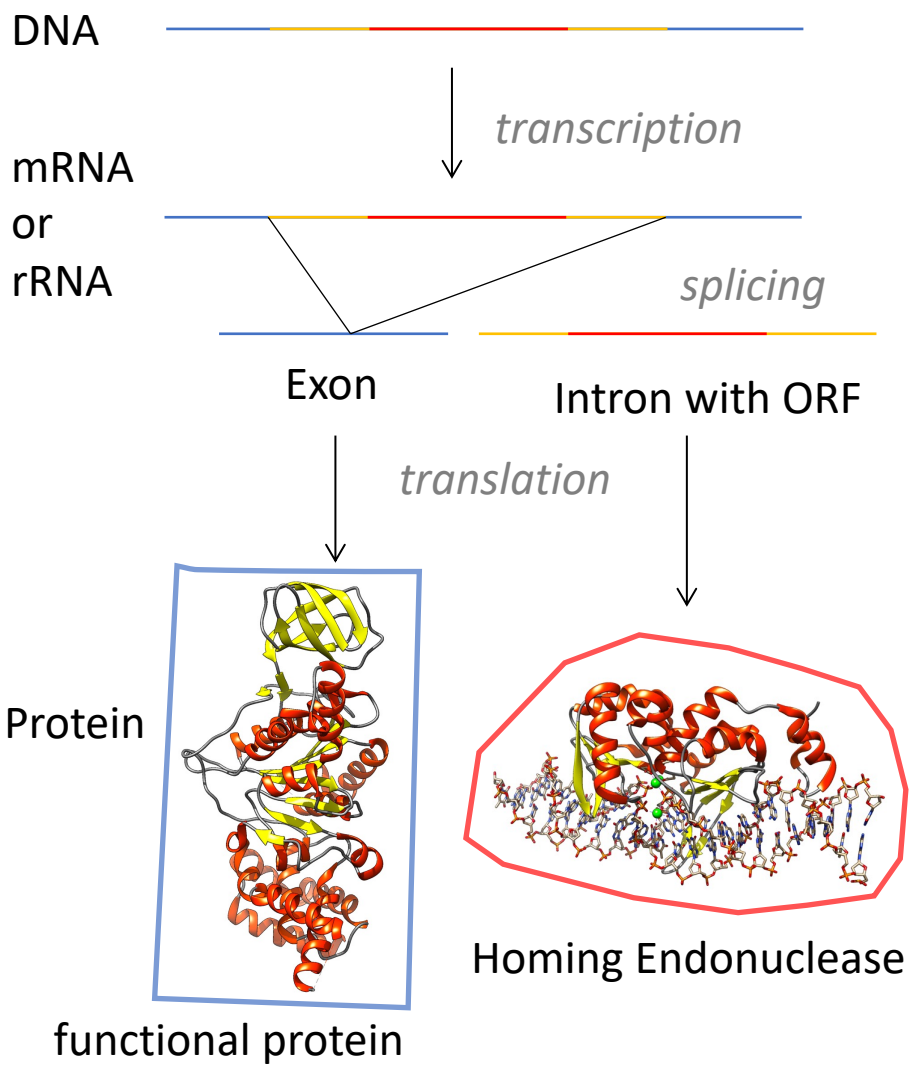
FIG. 8. **Northern blot analysis of the *VMA1* mRNA.** Poly(A)$^+$ RNA was isolated from ANY21, size fractionated on an agarose-formamide gel (1.2% gel), and transferred to a nylon membrane filter. Blots were hybridized with the DNA probes shown in *A*. *Hatched region* represents the nonhomologous insert (see "Results"). Hybridization patterns for probe-1 (*lane 1*) and probe-2 (*lane 2*) are shown in *B*. About 5 μg of RNA was loaded in each *lane*.

Levels of Selection



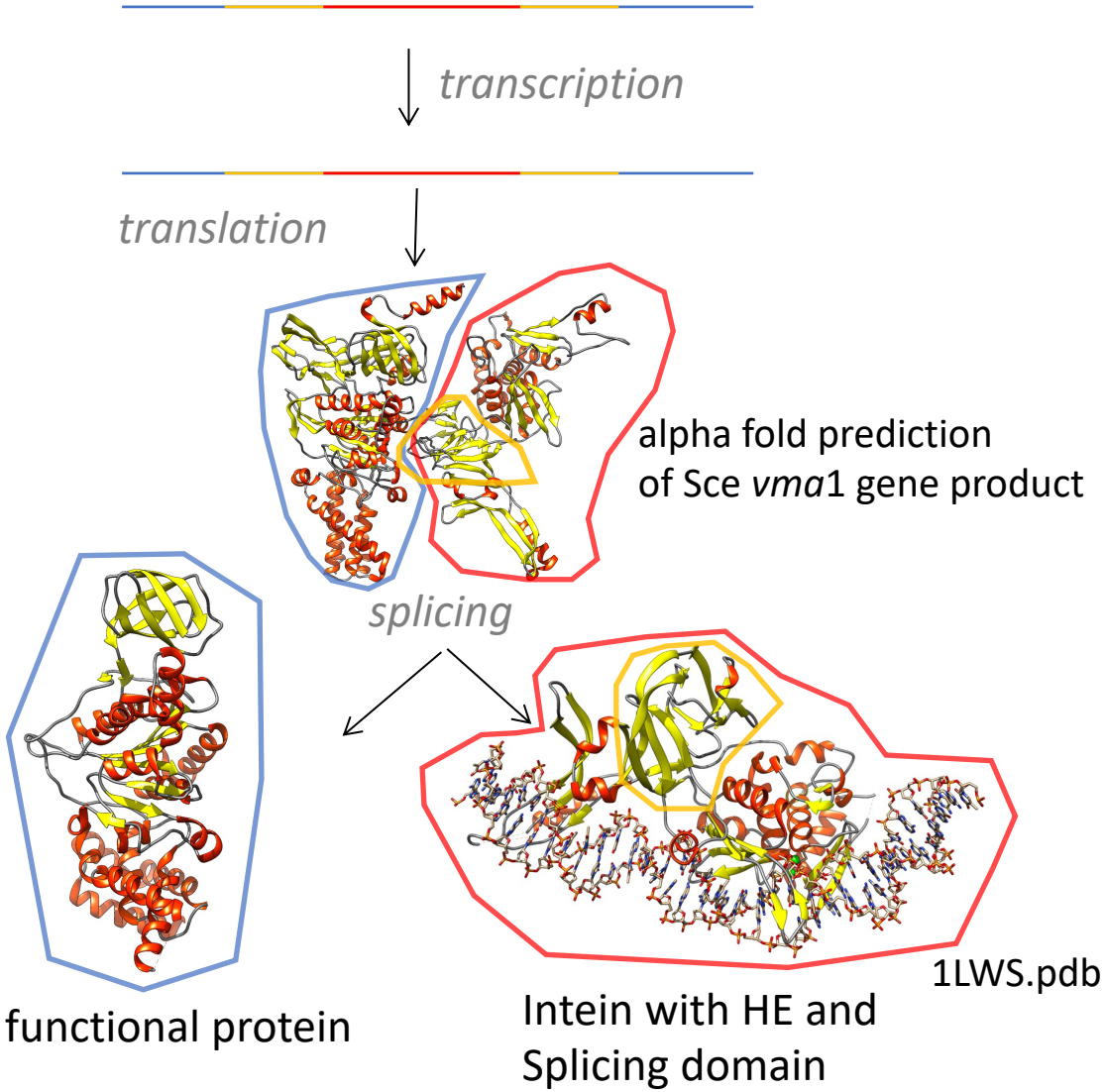
Group I Intron

Exon: Intron: ORF on Intron



Intein

Extein: Intein Splicing Domain: HE domain:

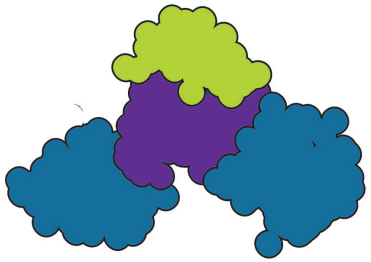


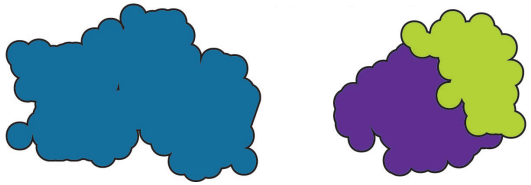
Inteins (molecular fleas): Self-Splicing Protein Mobile Elements

DNA

RNA

Protein





Intein Self-Splicing Domains

Auto catalytic splicing reaction removes the intein from the host protein

Intein Homing Endonuclease Domain

Recognizes unoccupied intein insertion sites (IIS) and disrupts the sequence with a double strand break

Extein

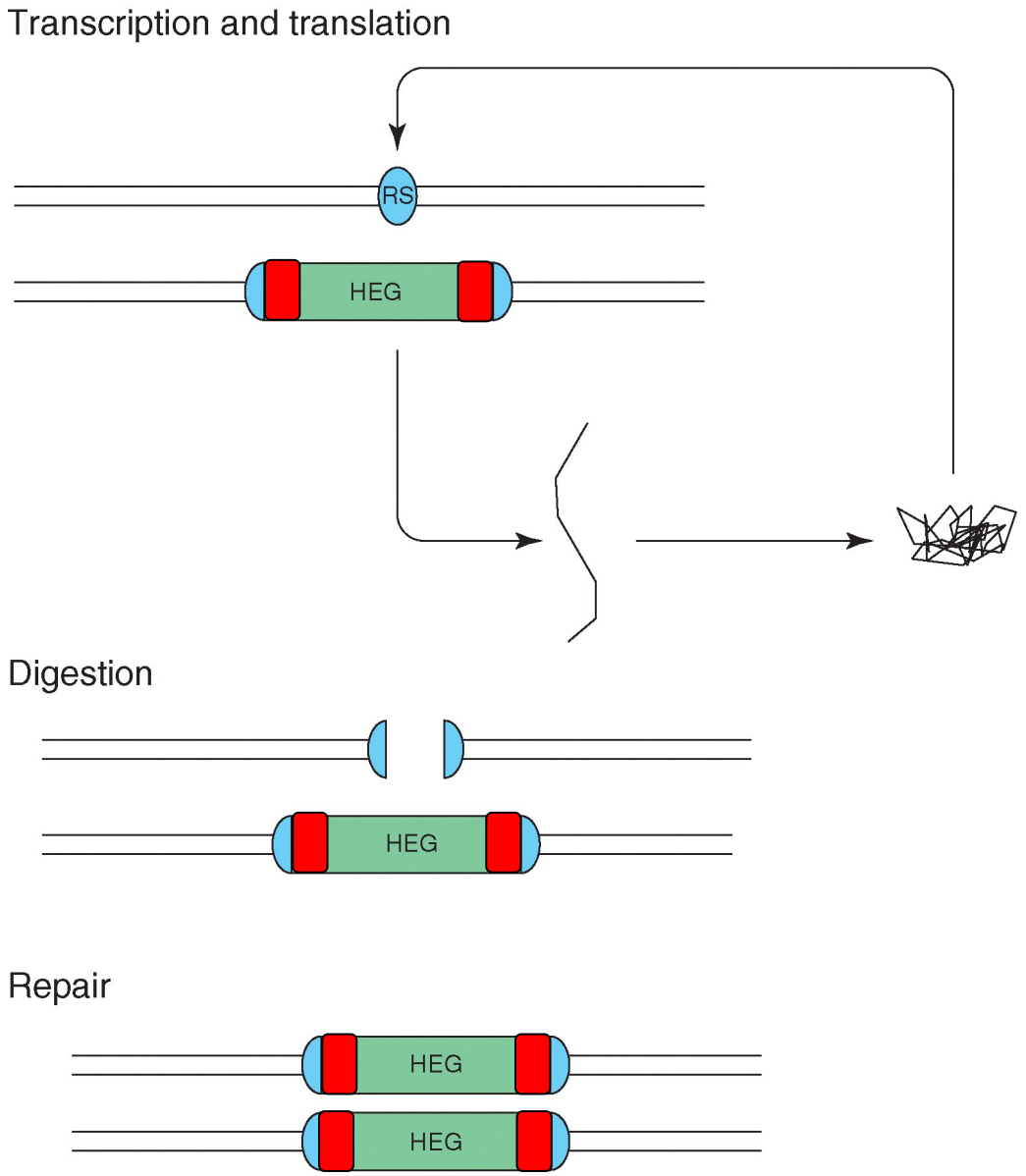
Host protein

Intein Homing

Intein = **Int**ervening Prot**ein** sequence
= “protein intron”

Self Splicing Domain

Homing Endonuclaease Domain



Modified from

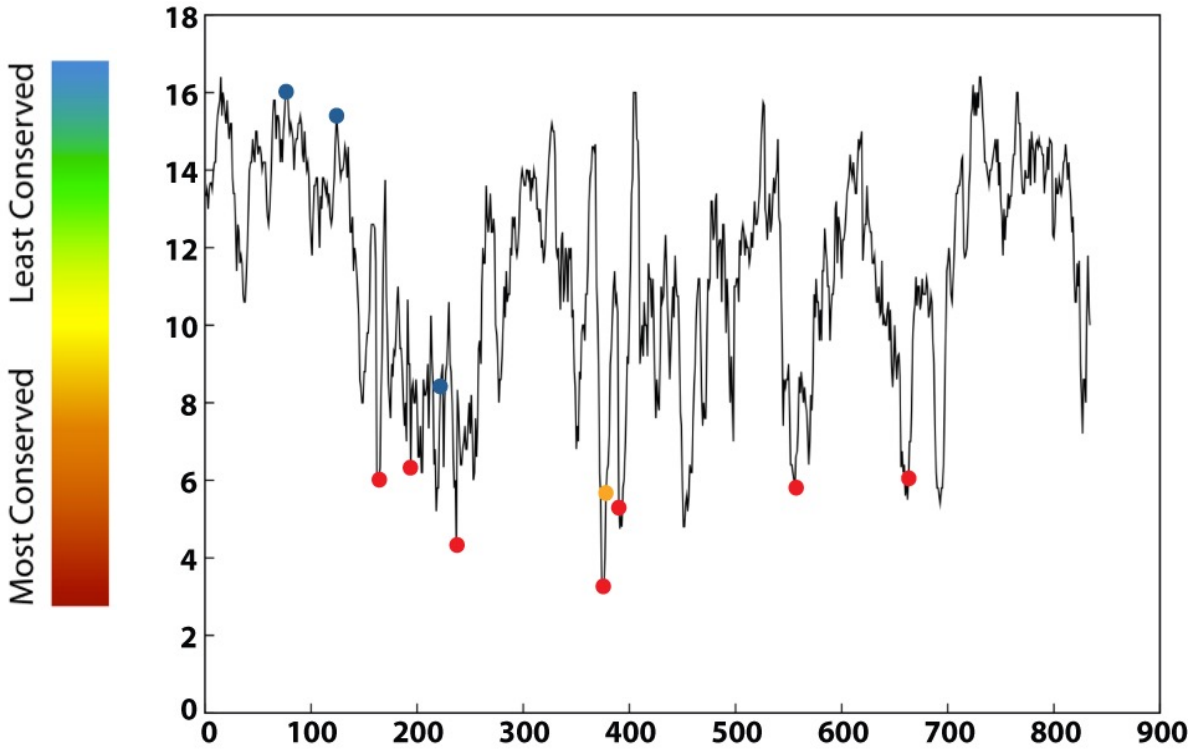
Homing endonuclease genes: the rise and fall and rise again of a selfish element

Austin Burt and Vassiliki Koufopanou

Current Opinion in Genetics & Development

Volume 14, Issue 6, December 2004, Pages 609-615

Ribonucleotide Reductase



Intein
Group I Introns
Group II Introns

Swithers KS, Senejani AG, Fournier GP, Gogarten JP (2009)
Conservation of intron and intein insertion sites: Implications for life histories of parasitic genetic elements.
BMC Evolutionary Biology 2009, 9:303 doi:10.1186/1471-2148-9-303 (Highly Accessed)

In case of inteins the co-evolution of genes results in

- **Mutualism** between splicing and homing endonuclease domain

- **Commensalism** between host protein and intein without HE domain

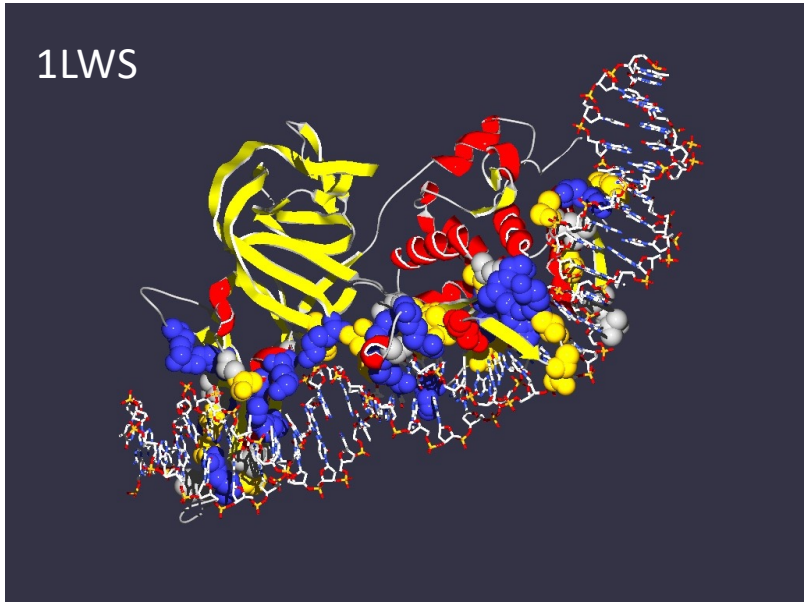
- **Parasitism** between host protein and intein with HE

Encoding

Extein = Host Protein

Intein Splicing Domain

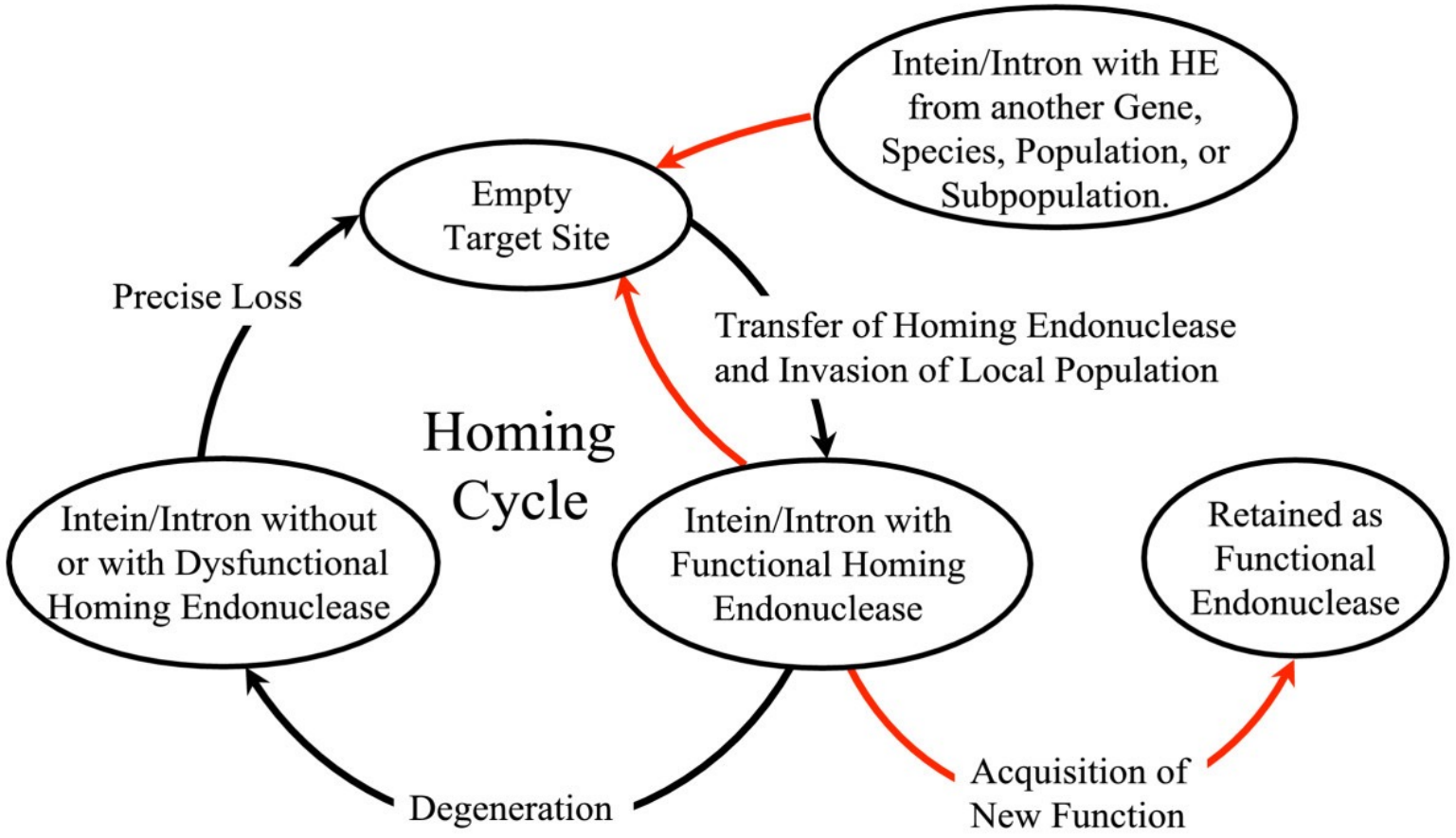
Homing Endonuclease (HE) domain



Structure from Moure et al. (2002) Nat.Struct.Biol. 9: 764-770

How can inteins with functional homing endonuclease survive on the long run?

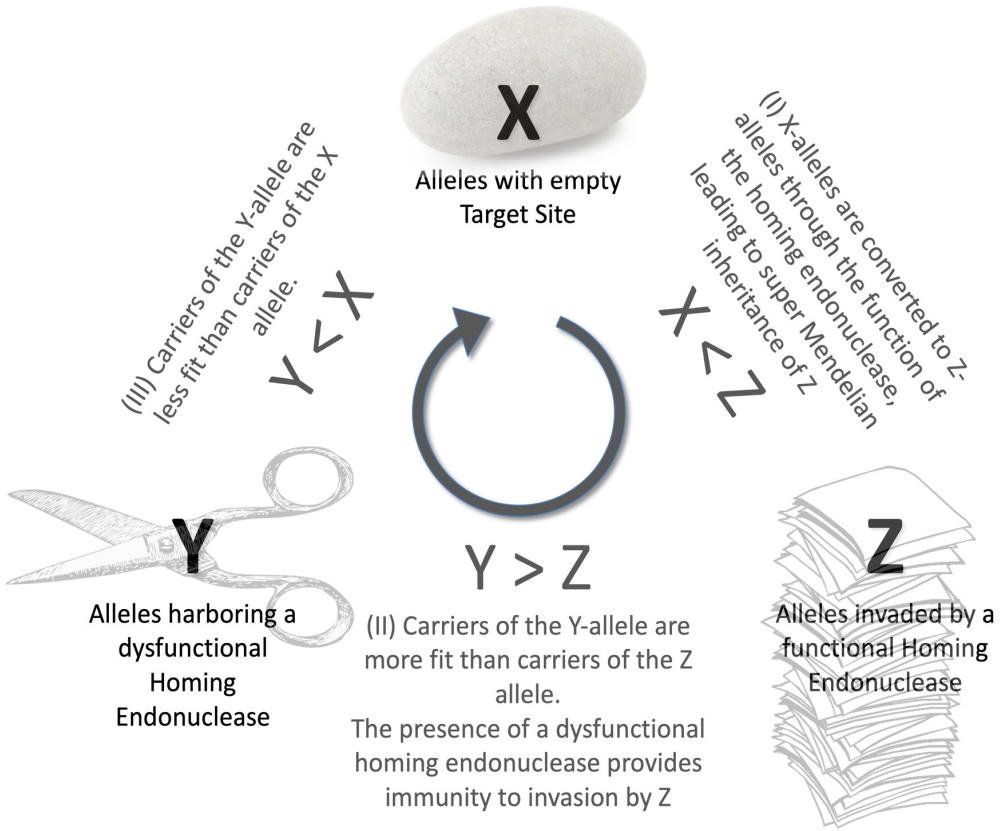
- Homing cycle
- Coexistence of the three forms (preditor prey with three partners, in an intransitive fitness relationship)
- Diverse environment with regions that select for the intein, and those that don’t

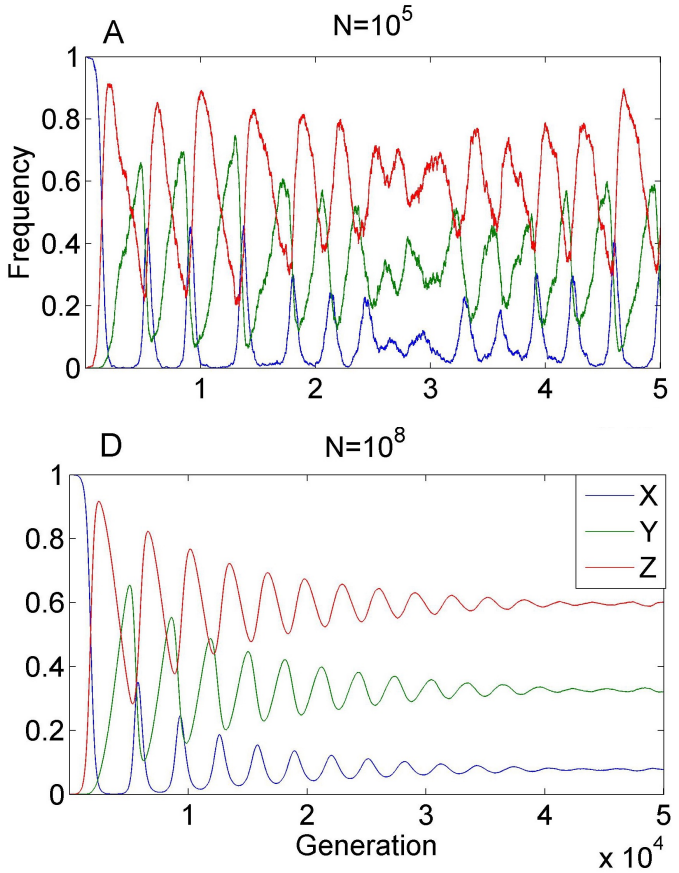


Homing cycle of a parasitic genetic element (modified from [3, 13]). Recent findings suggest that due to complex population structure the cycle might not operate in synchrony in different subpopulations. The red arrows indicate the trajectory of the functioning HE and the black arrows the fate of the host gene. The precise loss can occur through recombination with an intein or intron free allele, or, in case of introns, through recombination with a reverse transcript of the spliced mRNA [39, 40].

Gogarten, J. Peter and Elena Hilario (2006)
Inteins, introns, and homing endonucleases: Recent revelations about the life cycle of parasitic genetic elements.
BMC Evolutionary Biology 6:94 Open Access (Highly Accessed)

Coexistence of the three forms





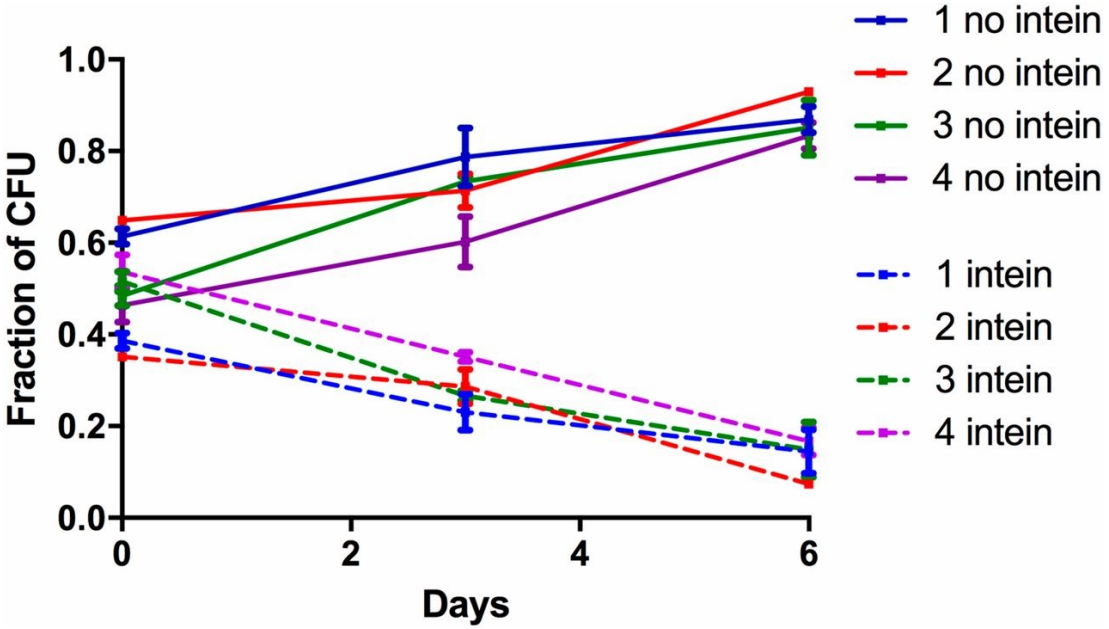
A Molecular Rock-Paper-Scissors game

Barzel *et al. BMC Evolutionary Biology* 2011 **11**:324 doi:10.1186/1471-2148-11-324

Simulations using difference-equations in populations of limited size

Inteins have a high fitness cost for the host organism.

United States – Israel
Binational Science Foundation



Competition Experiment between Intein + and Intein – strains (otherwise isogenic)

From: Impact of a homing intein on recombination frequency and organismal fitness. Naor et al.
doi: 10.1073/pnas.1606416113

Dr. Adit Naor
Tel Aviv Univ / Stanford

Dr. Shannon Soucy
UConn / Dartmouth

Dr. Uri Gophna
Tel Aviv Univ

Dr. Thane Papke
UConn

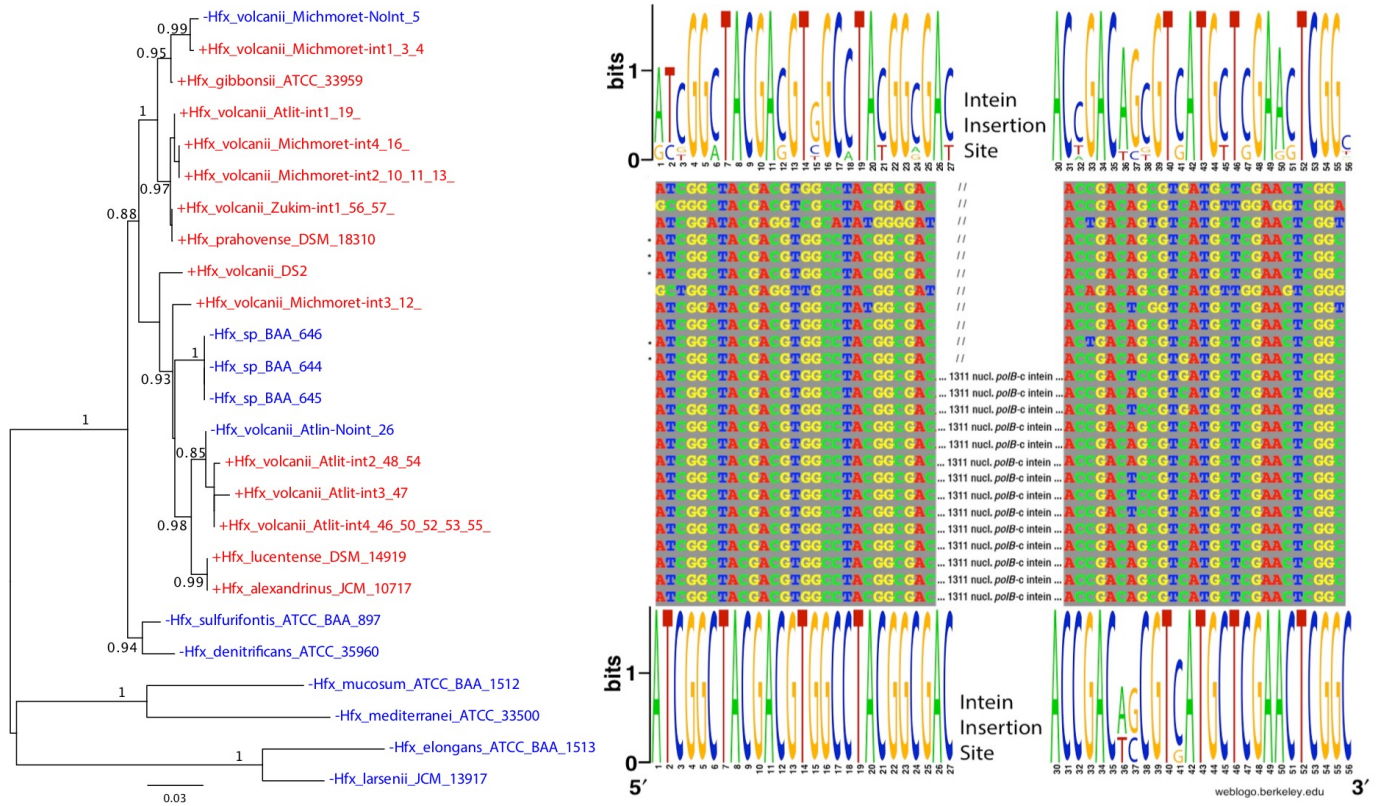
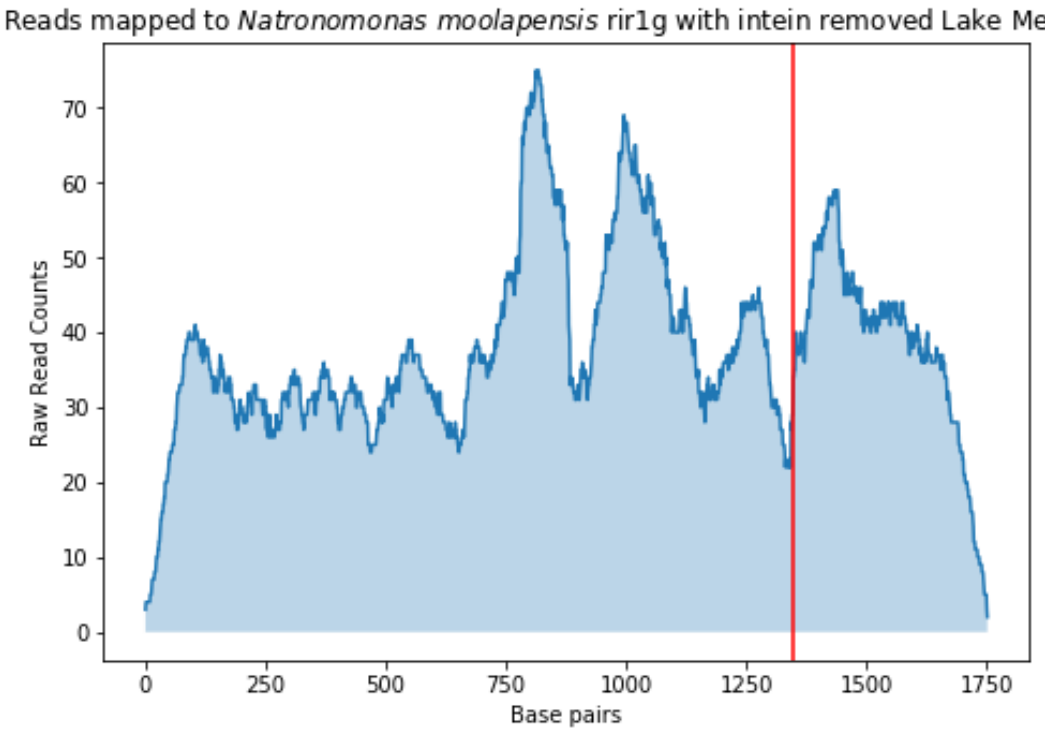
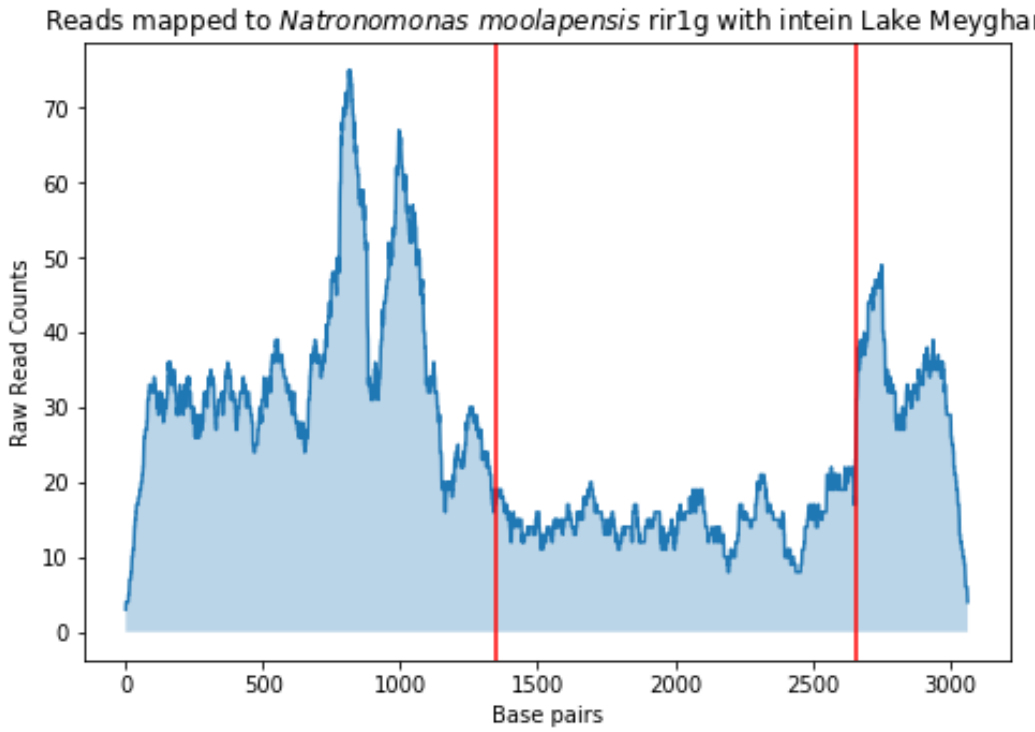


Fig 6. Maximum likelihood phylogeny for *polB* extein sequences (left) and conservation of *polB*-c intein insertion sites (right). Numbers give support values calculated using the approximate Likelihood Ratio Test as implemented in phyml 3.0 (32). Although drawn as a rooted, the tree should be considered unrooted. The finding that sequences without (blue) and with intein (red) do not always form distinct clans (34) reveals that invasion of the *Haloferax* genus with the *polB*-c intein is an ongoing process. The panel on the right shows a *polB* nucleotide sequence alignment around the intein insertion site c. Web logos (33) give the site conservation for intein minus (top) and intein plus sequences (bottom). The five intein minus sequences that group within the cluster of intein plus sequences are marked with an asterisk. The intein minus sequences show greater nucleotide diversity surrounding the intein insertion site, mainly in synonymous positions -- only two positions at the 5' and close to the 3' end of the alignment represent non-synonymous changes. Homing endonuclease site specificity was shown to tolerate substitutions that result in non-synonymous changes (35), suggesting that none of the depicted *Haloferax* sequences may be immune to intein invasion.

Frequency of reads mapped to *Natronomonas moolapensis* rir1g

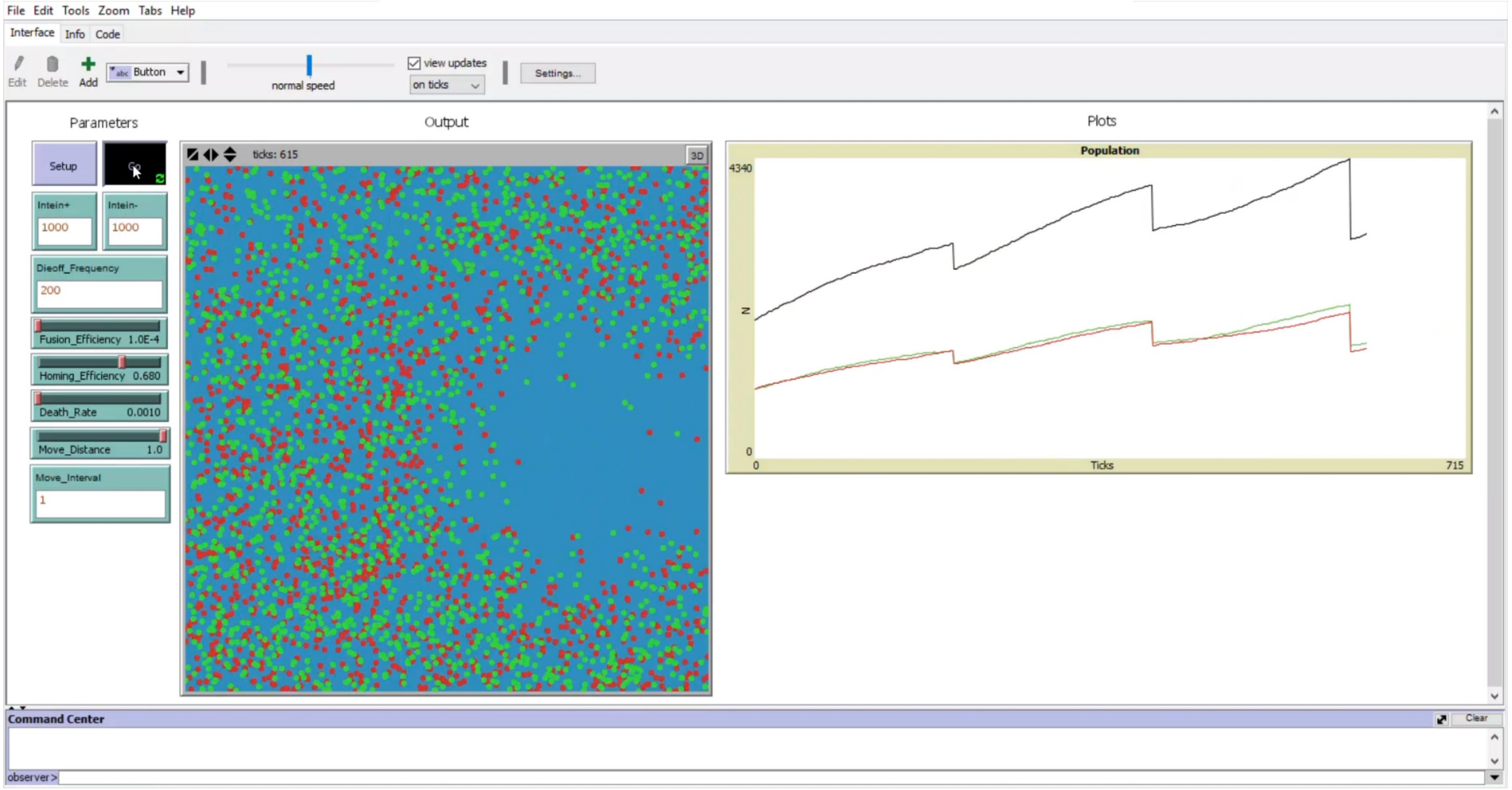




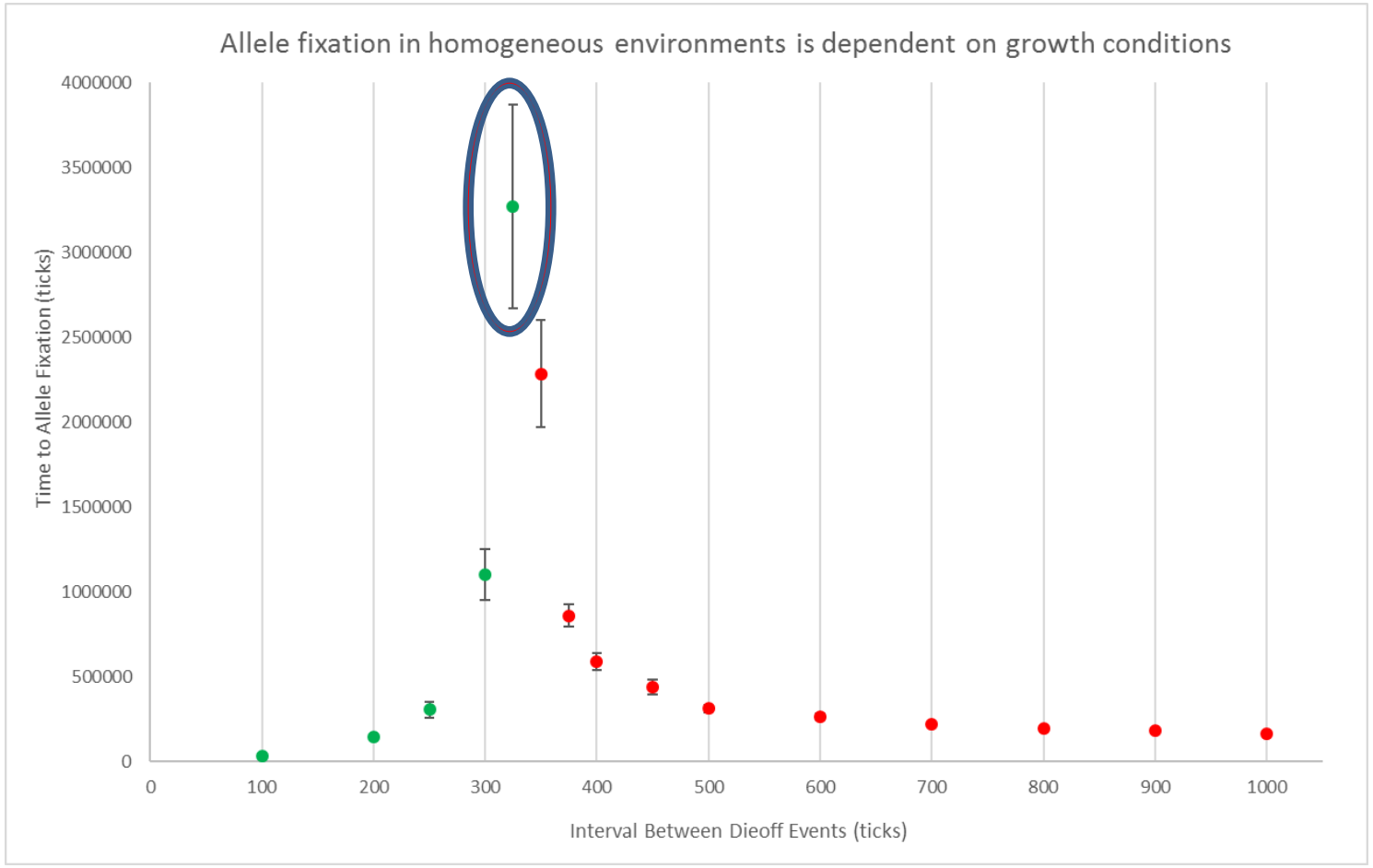
- Metagenomic reads from Lake Meyghan* mapped back to reference sequences with and without the intein.
- Intein was artificially removed, to intact insertion site.
- Red lines indicate intein boundaries, sharp decrease in coverage when intein removed.

*Naghoni, A. *et al.* Microbial diversity in the hypersaline Lake Meyghan, Iran. *Sci. Rep.* **7,** (2017).

Simulation: Homogeneous Environment with localized random extinctions

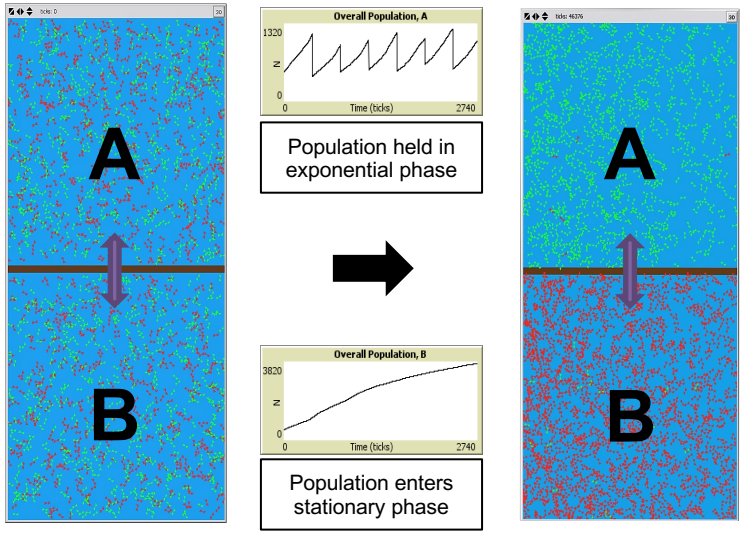


Agent based modeling using NetLogo

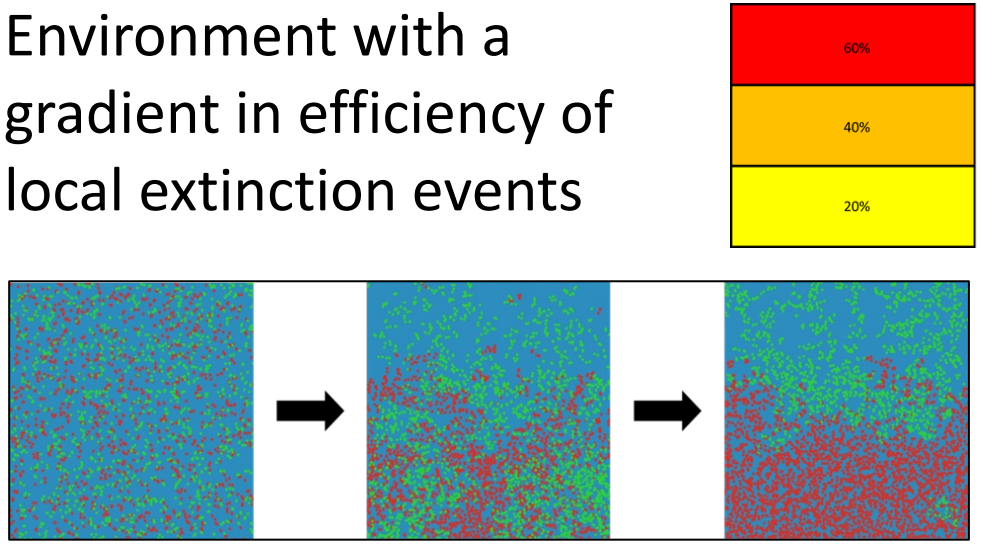

Allele fixation in homogeneous environments is dependent on growth conditions
Time to Allele Fixation (ticks)
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
0
100
200
300
400
500
600
700
800
900
1000
Interval Between Dieoff Events (ticks)

Changing conditions (temporal and/or spatial) can facilitate the long-term coexistence of homing endonucleases with empty target sites

Two environments with different frequencies of local extinction events



Environment with a gradient in efficiency of local extinction events

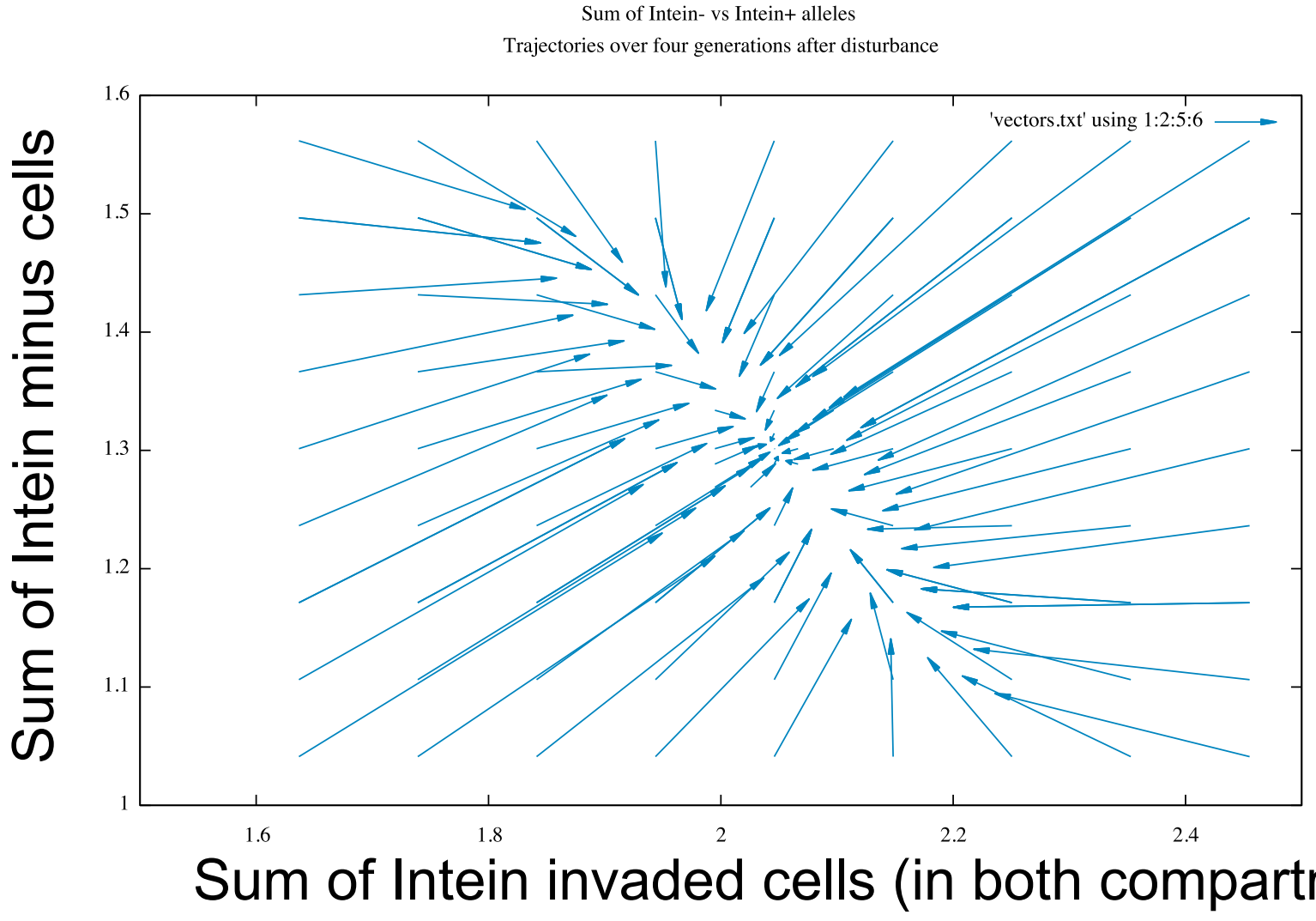


Close to stationary phase: Inteins transmission continues, increasing the frequency of the invaded allele

During active growth: Due to fitness cost of the intein, the frequency of the noninvaded allele increases

A heterogeneous environment can lead to long-term stable coexistence of invaded and uninvaded genes

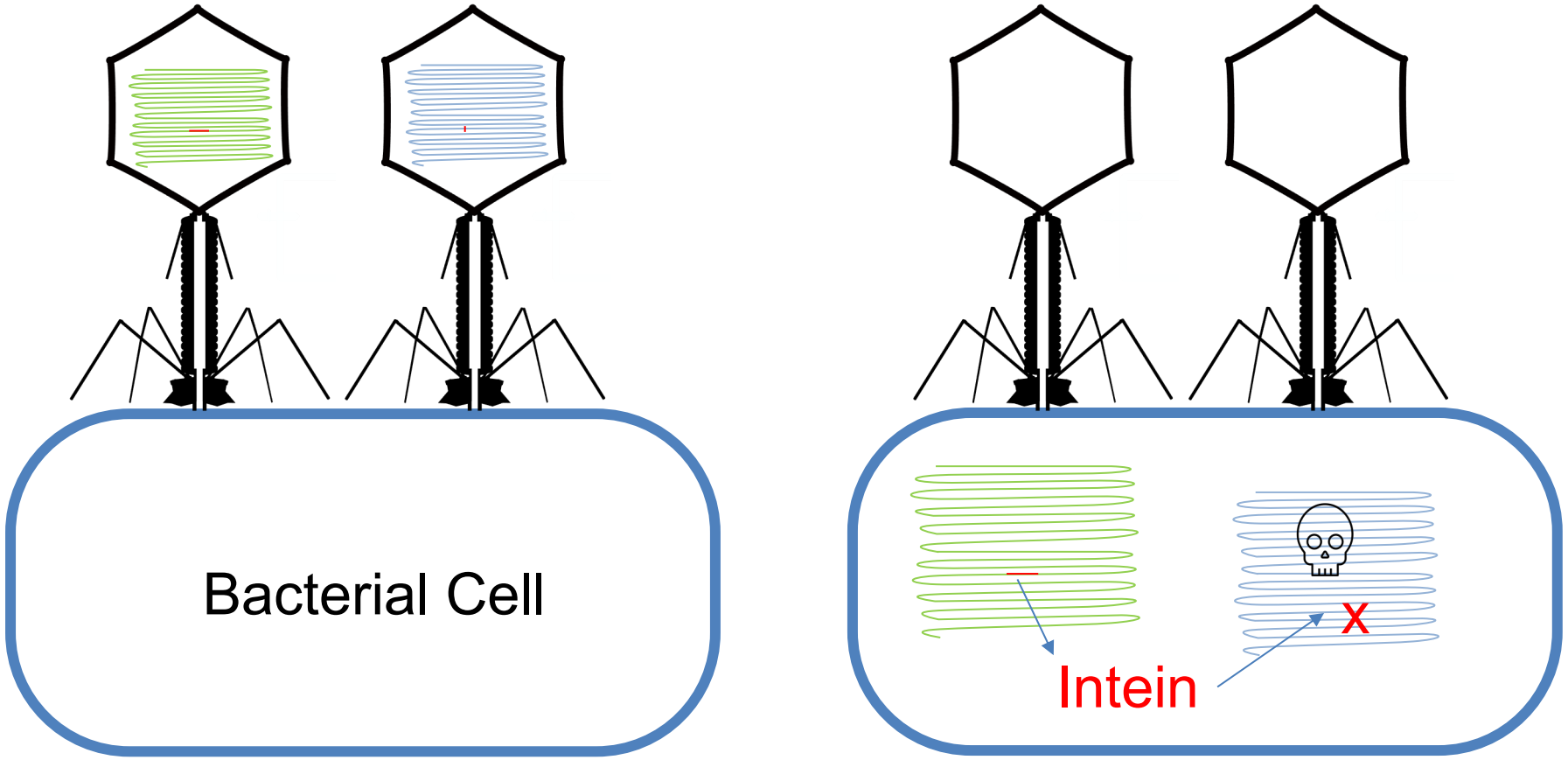
Simulation of a two-compartment system (each of the compartments is homogenous) using iterations.
In each discrete generation 10% of the population dies in compartment 2, and .1% move between compartments. Without death the carrying capacity is 2 in each compartment.



How can inteins with functional homing endonuclease survive on the long run?

- Homing cycle
- Coexistence of the three forms (preditor prey with three partners, in an intransitive fitness relationship)
- Diverse environment with regions that select for the intein, and those that don’t
- They picked up a function that increases the fitness of the host (emergency shut off, if conditions are bad – e.g., salt, temperature, redox potential; mating type switching endonuclease, weapon against competitors).

Phage with Intein Phage w/o Intein

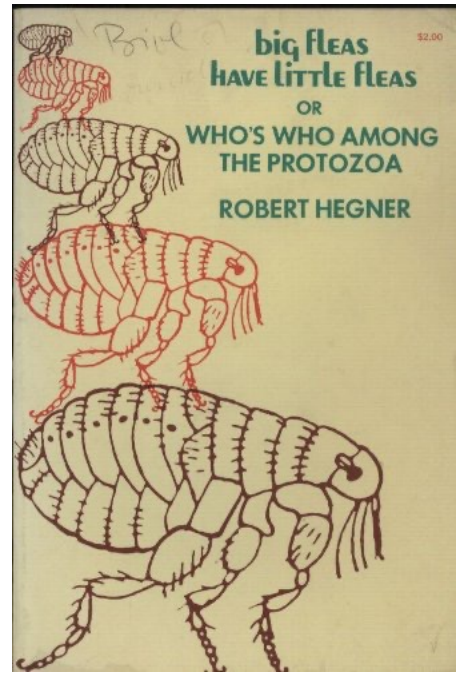


Infinite Regress

After a lecture on the structure of the solar system, the cosmologist William James was accosted by a little old lady.
"Your theory that the sun is the centre of the solar system, and the earth is a ball which rotates around it has a very convincing ring to it, Mr. James, but it's wrong. I've got a better theory," said the little old lady.
"And what is that, madam?" inquired James politely.
"**That we live on a crust of earth which is on the back of a giant turtle**."
Not wishing to demolish this absurd little theory by bringing to bear the masses of scientific evidence he had at his command, James decided to gently dissuade his opponent by making her see some of the inadequacies of her position.
"If your theory is correct, madam," he asked, "**what does this turtle stand on?**"
"You're a very clever man, Mr. James, and that's a very good question," replied the little old lady, "but I have an answer to it. And it's this: **The first turtle stands on the back of a second, far larger, turtle, who stands directly under him**."
"**But what does this second turtle stand on?**" persisted James patiently.
To this, the little old lady crowed triumphantly,
"**It's no use, Mr. James—it's turtles all the way down**."

From *Ross, John R. (1967). Constraints on variables in syntax (Doctoral dissertation). MIT. hdl:1721.1/15166.*
via https://en.wikipedia.org/wiki/Turtles_all_the_way_down

Siphonaptera

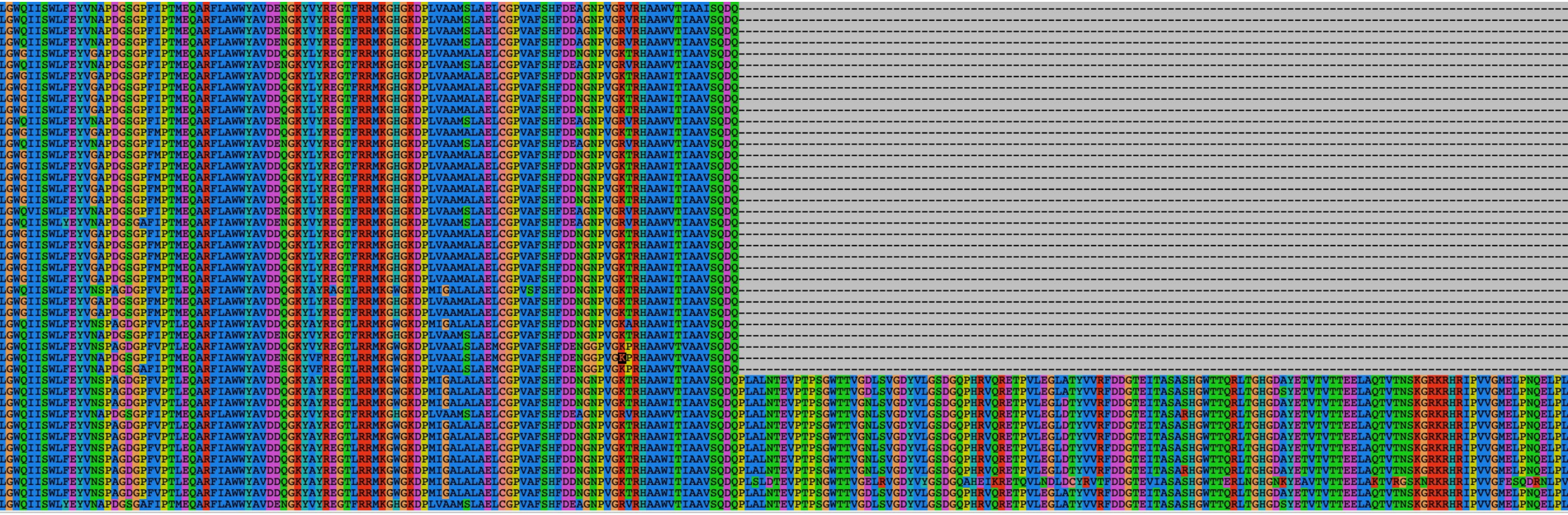


Great fleas have little fleas upon their backs to bite 'em,
And little fleas have lesser fleas, and so *ad infinitum*.
And the great fleas themselves, in turn, have greater fleas to go on ;
While these again have greater still, and greater still, and so on.

From: Augustus De Morgan's poem *Siphonaptera* (1872)
via https://en.wikipedia.org/wiki/Siphonaptera_(poem)

Terminase Subunits from Actinobacteriophages Cluster A1

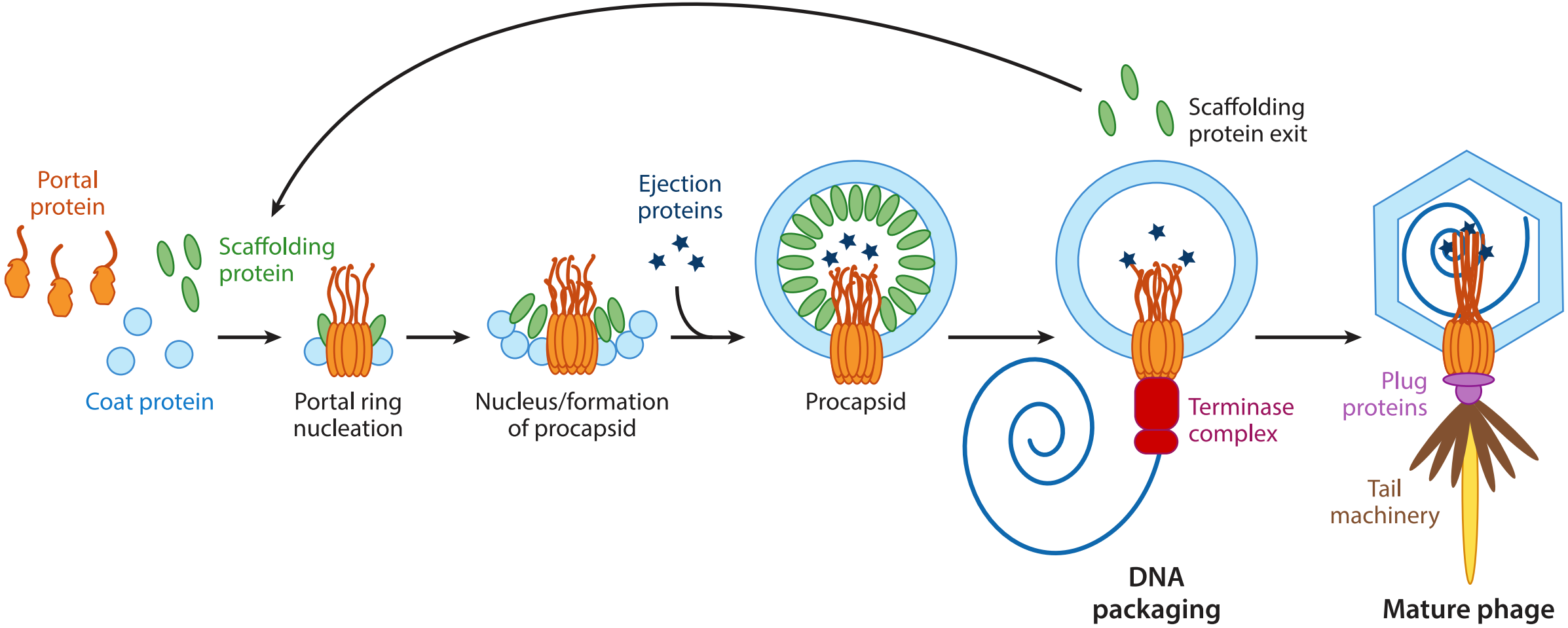
Alignment of protein sequences



N-Extein

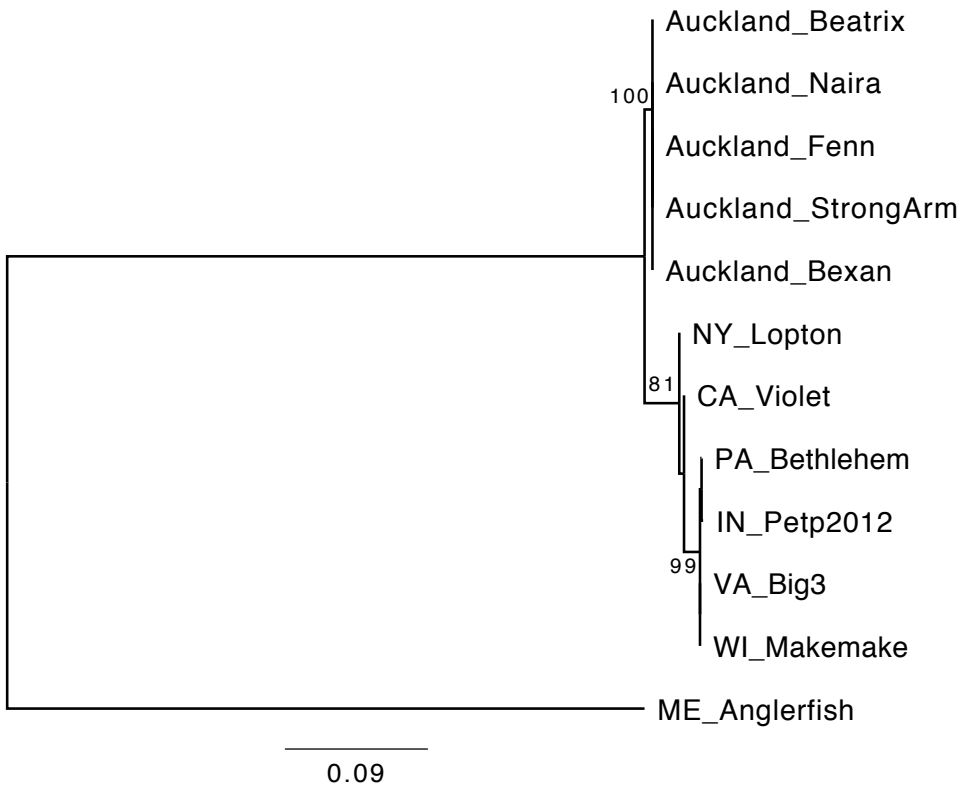
Intein

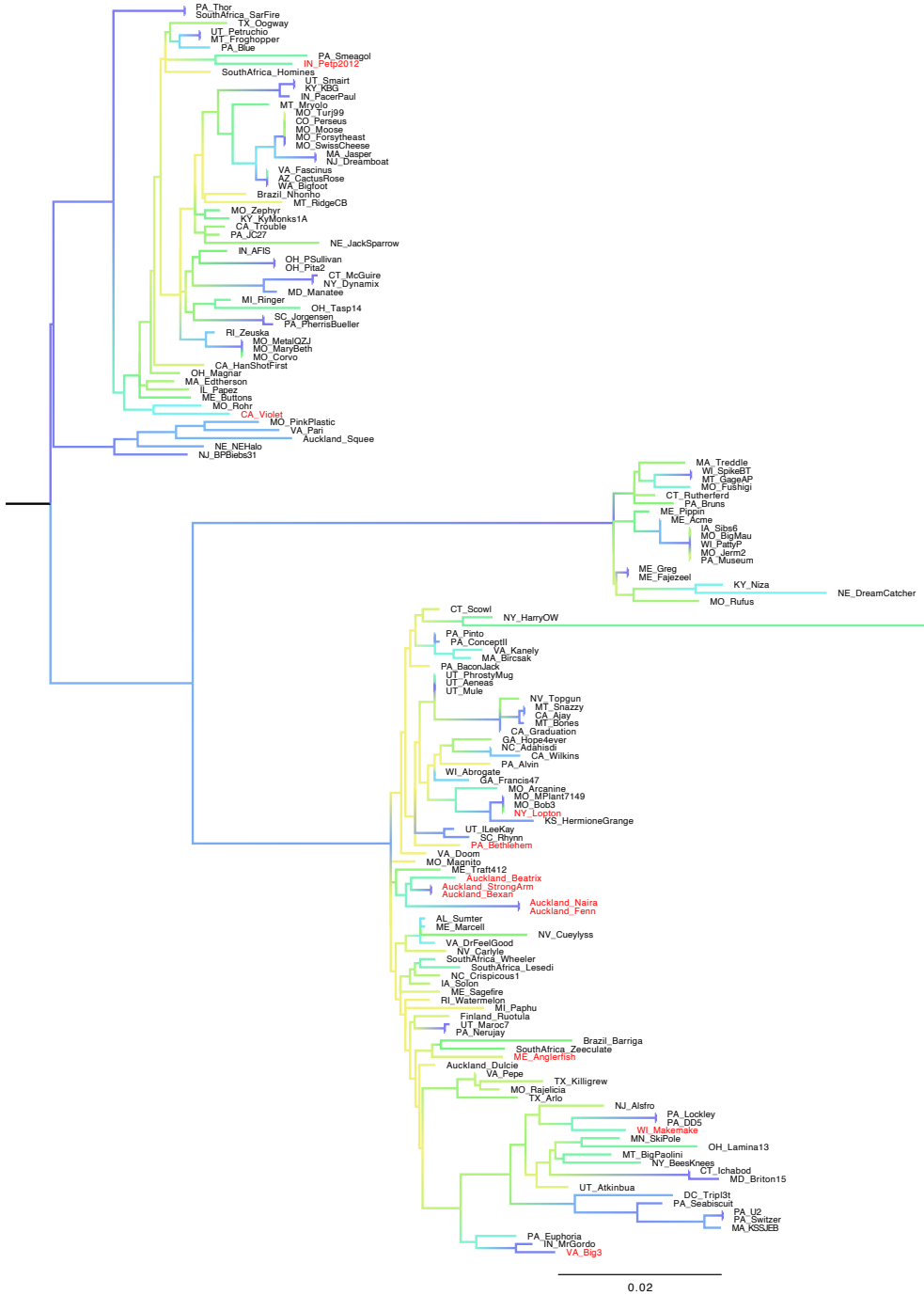
Phage Assembly



From: Corynne L. Dedeo, Gino Cingolani, and Carolyn M. Teschke:
Portal Protein: The Orchestrator of Capsid Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses
Annual Review of Virology 2019

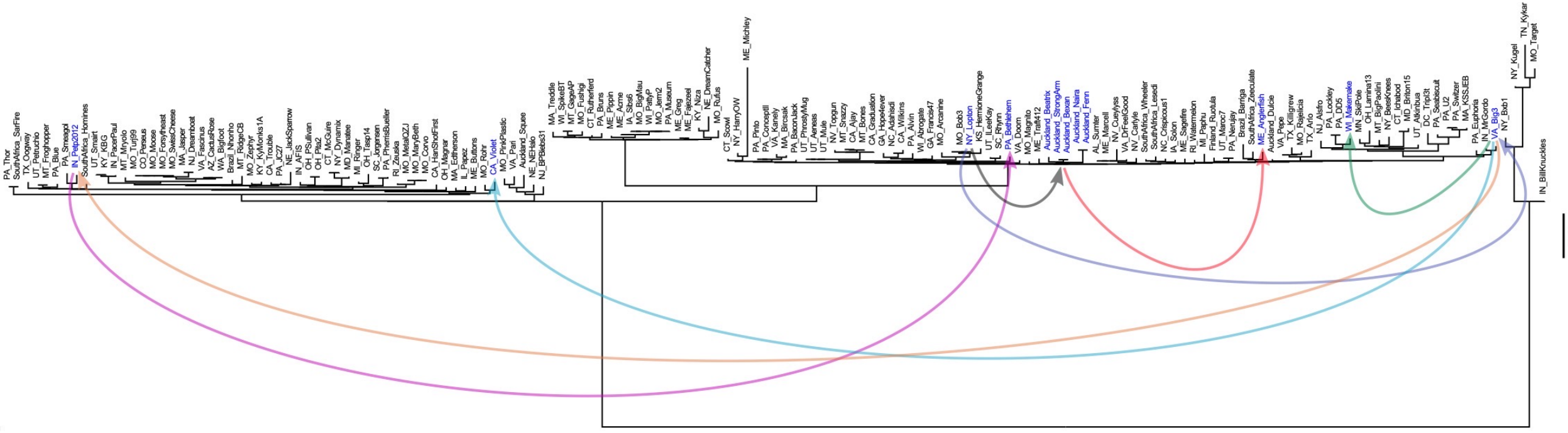
Exteins and their inteins from homologous terminase subunits in the A1 cluster





Extein

Reconciliation between extein and intein trees



This is one of many equally parsimonious reconciliations (allowing for transfer, duplication and loss).

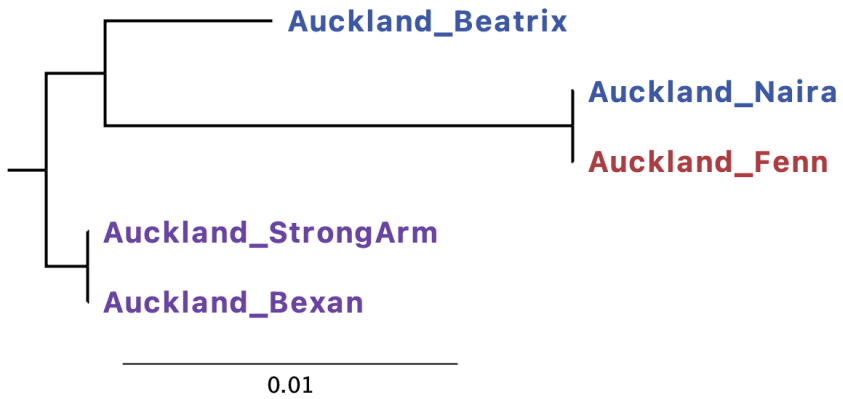
Note that the Auckland cluster requires only one intein acquisition at the base.

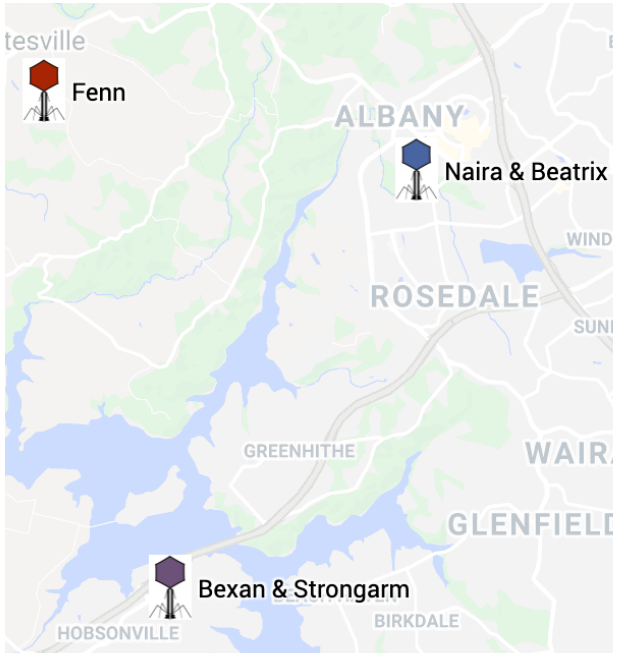
The A1 Phages from the Auckland area

Name	Year of Isolation	Location	Description of Location
Fenn	2017	36.724847 S, 174.635248 E	Soil on an old metal garage door in a field near Auckland
Naira	2017	36.734195 S, 174.700783 E	Massey University community garden compost bin
Beatrix	2017	36.734195 S, 174.700783 E	Massey University community garden compost bin
Bexan	2019	36.790857 S, 174.65929 E	Catalina community garden compost bin
Strongarm	2019	36.790857 S, 174.65929 E	Catalina community garden compost bin

	Number of sites	Number of sequences	Number of Polymorphic sites (only SNPs)	Watterson's Theta
Extein	1689	5	31	0.0088
Intein	1023	5	0	0

Phylogeny based on extein sequences (topology is the same if the whole genomes are used):





The Evolutionary History of a DNA Methylase Reveals Frequent Horizontal Transfer and Within-Gene Recombination

by Sophia P. Gosselin [1], Danielle R. Arsenault [1], Catherine A. Jennings [1] and Johann Peter Gogarten [1,2,*]

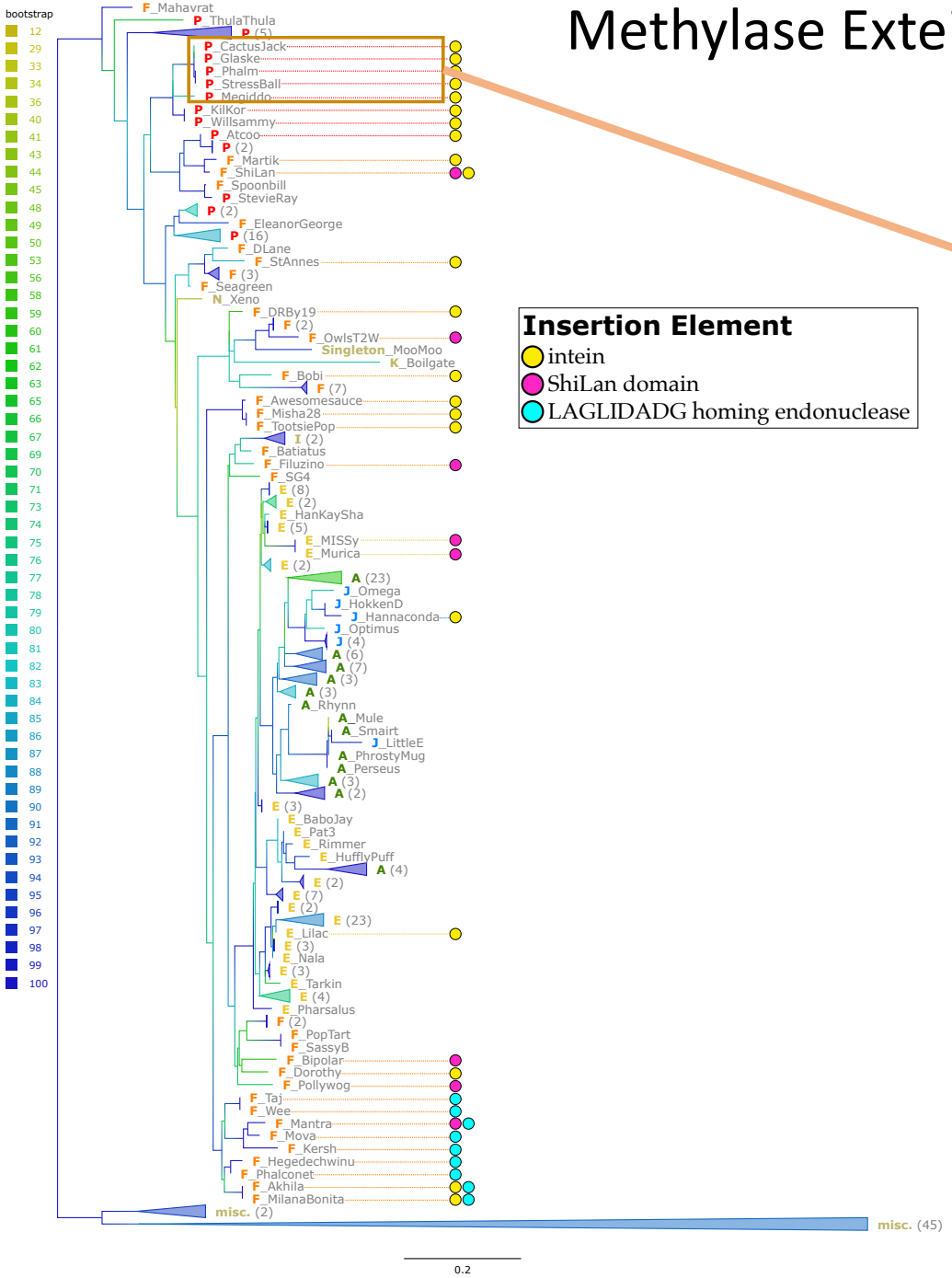
[1] Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT 06268-3125, USA

[2] Institute for Systems Genomics, University of Connecticut, Storrs, CT 06268-3125, USA

[*] Author to whom correspondence should be addressed.

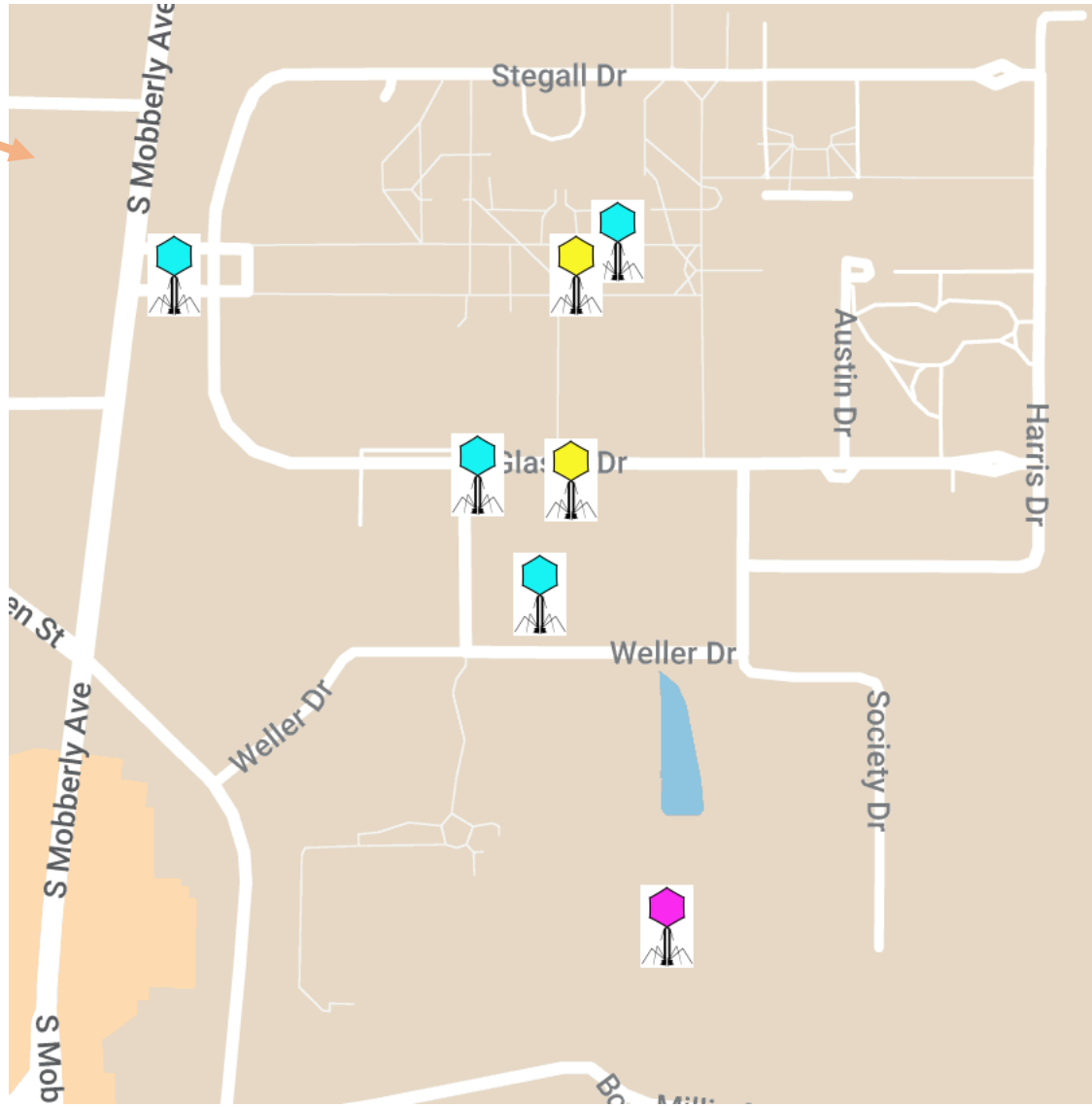
Genes **2023**, *14*(2), 288; https://doi.org/10.3390/genes14020288

Methylase Extein Phylogeny



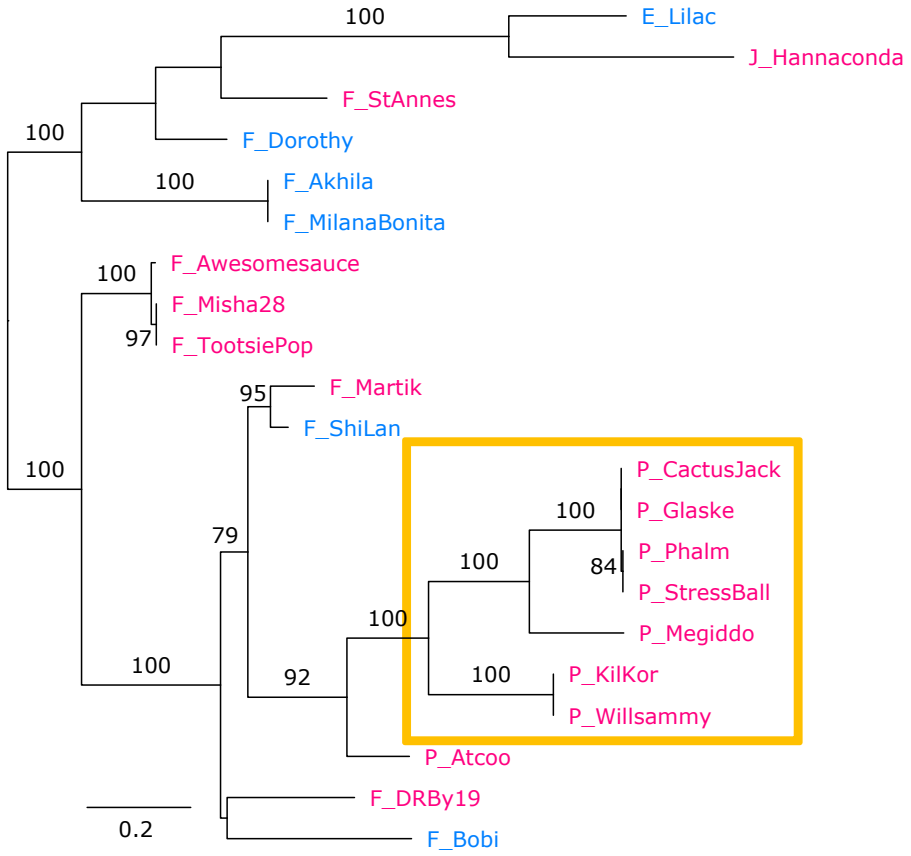
Insertion Element

- intein
- ShiLan domain
- LAGLIDADG homing endonuclease



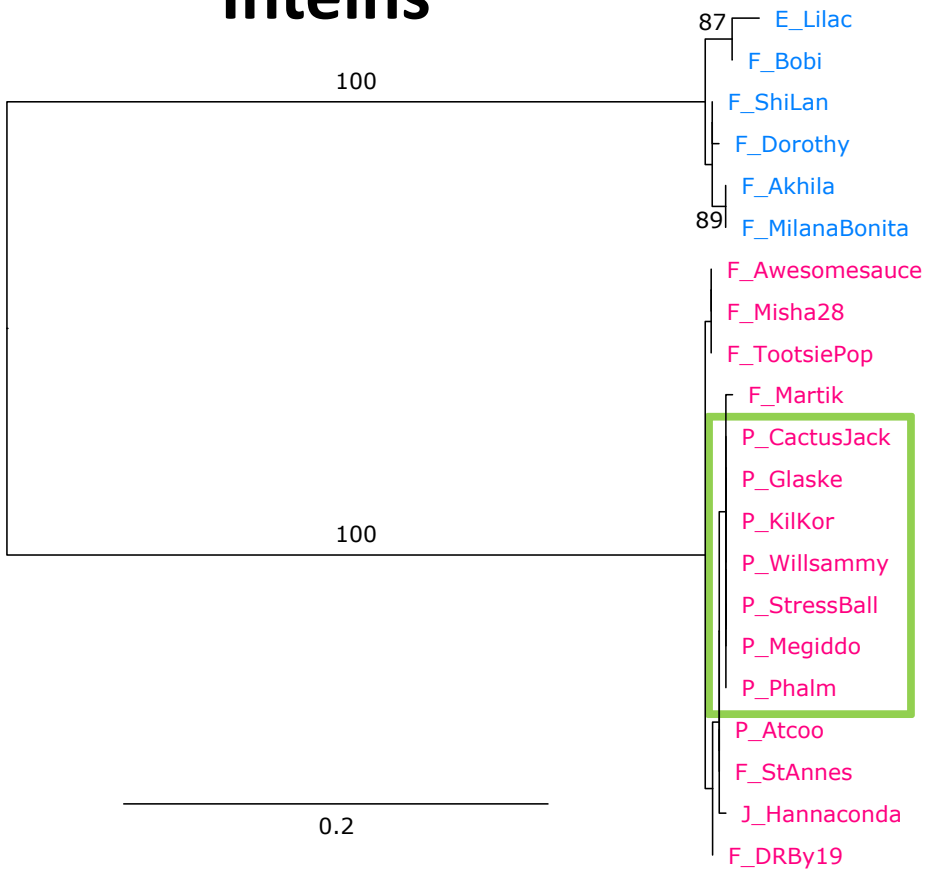
Phylogeny of an Actinophage Methylase Family

Exteins



Max **13%** sequence divergence

Inteins



Max **0%** sequence divergence