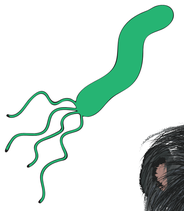
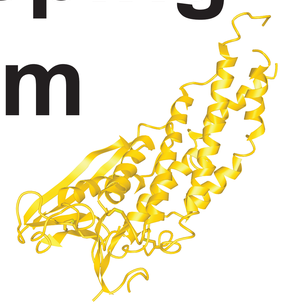
Rapid "mix-n-match" evolution of a housekeeping protein in response to bacterial antagonism



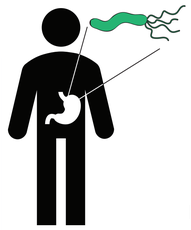
EmilyClare Baker [1], Kristin Kohler[1], Matt Barber[1]

[1]University of Oregon Institute of Ecology and Evolution



Helicobacter pylori

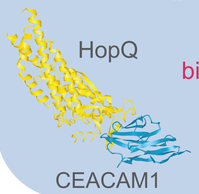
A ubiqutous, pathogenic gut microbe[1]



- Colonizes ~50% of all humans
- Infection with *H. pylori* can cause...
  - o Gastritis
  - o Peptic ulcers
  - o Gastric cancer*

*The WHO classifies *H. pylori* as a type I carcinogen[2]

H. pylori encodes the CEACAM binding adhesin HopQ[3]

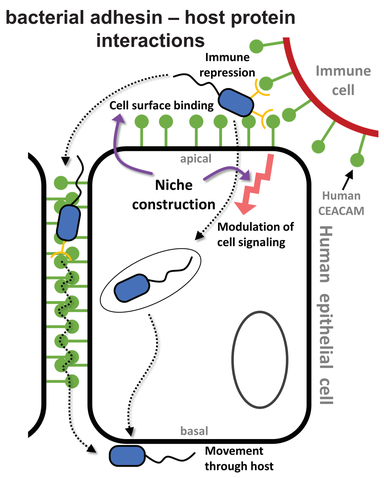


CEACAM1 - HopQ binding is important for *H. pylori* niche construction and virulence.

Bacterial adhesin proteins can exploit cell surface proteins for...[4]

- niche construction
- immune evasion
- movement through the host

Potentially decreasing host fitness

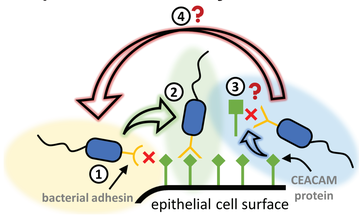


Consequences of evolutionary conflict between pathogens and host 'housekeeping genes'

- Pathogen antagonism is predicted to lead to an 'evolutionary arms race'
  - characterized by rapid evolution host and bacterial proteins.

How host 'housekeeping' proteins can evolve to evade pathogen exploitation and maintain essential functions is an open question.

Model of bacterial adhesin host protein evolutionary arms race



1. Bacterial adhesin is unable to bind to the host epithelial protein
2. Pressure to manipulate host drives evolution of bacterial adhesin to bind host protein
3. Pressure to prevent exploitation favors evolution of host protein away from adhesin binding
4. Pressure to evolve shifts back to the bacterial adhesin

If and how housekeeping genes can evolve against pathogen exploitation is unclear.

The CEACAM protein family[4]

A multifunctional cell adhesion family in vertebrates

Functions include...

- Cell adhesion
- Intra- and Intercellular signaling
- Immune function
- Cell cycle modulation

Several bound by bacteria

- CEACAM1
- CEACAM5 (CEA)
- CEACAM6
- CEACAM3

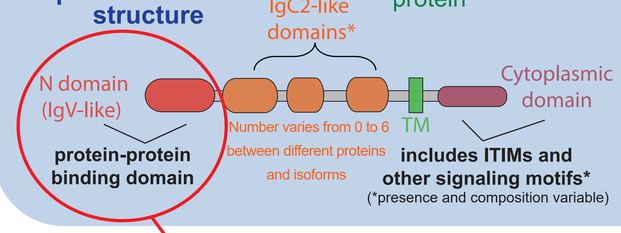
CEACAMs 1, 5 & 6

- housekeeping proteins involved in cellular signaling
- expressed on epithelia (esp. gastric) and immune cells
- strong markers of cancer progression and prognosis

CEACAM3

- expressed on immune cells
- likely a dedicated immune protein

CEACAM protein domain structure



Region bound by bacterial adhesins

CEACAMs bound by bacterial adhesins the most rapidly evolving in primates

Based on PAML[5] analysis of CEACAM proteins from 19 hominid, Old World Monkey and New World Monkey primates.

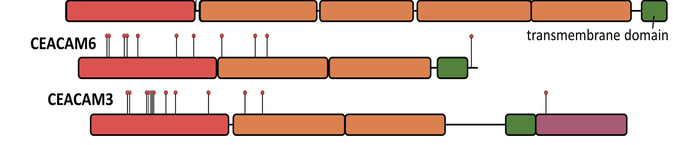
CEACAM gene	ℓ Model1	ℓ Model2	2δ	df	*P-value*	ℓ Model7	ℓ Model8	2δ	df	*P-value*
CEACAM1	-5299.42	-5244.56	109.71	2	7.50E-25	-5300.08	-5244.54	111.07	2	3.81E-25
CEACAM5	-5810.01	-5766.89	86.25	2	9.36E-20	-5811.24	-5766.93	88.61	2	2.86E-20
CEACAM3	-3484.59	-3445.21	78.75	2	3.96E-18	-3484.9	-3445.28	79.25	2	3.09E-18
CEACAM6	-3472.91	-3446.07	53.69	2	1.1E-12	-3472.98	-3446.13	53.7	2	1.09E-12
CEACAM8	-4878.78	-4855.3	46.97	2	3.15E-11	-4880.16	-4855.34	49.65	2	8.26E-12
CEACAM7	-2701.99	-2686.33	31.32	2	7.91E-08	-2702.21	-2686.41	31.61	2	6.84E-08
CEACAM20	-5435.75	-5430.42	10.66	2	0.00242	-5436.65	-5430.34	12.61	2	0.000913
CEACAM18	-2834.69	-2832.58	4.21	2	0.061	-2835.54	-2832.58	5.92	2	0.026
CEACAM19	-1860.75	-1860.04	1.41	2	0.246	-1860.77	-1860.04	1.45	2	0.242
CEACAM4	-1560.33	-1560.22	0.21	2	0.449	-1560.35	-1560.21	0.27	2	0.438
CEACAM21	-1606.24	-1606.17	0.15	2	0.463	-1606.26	-1606.17	0.18	2	0.457
CEACAM16	-3269.76	-3269.76	0	2	0.5	-3267.71	-3267.71	0	2	0.5

ℓ Model1 & ℓ Model7 log likelihood values for models without selection

ℓ Model2 & ℓ Model8 log likelihood values for models with selection

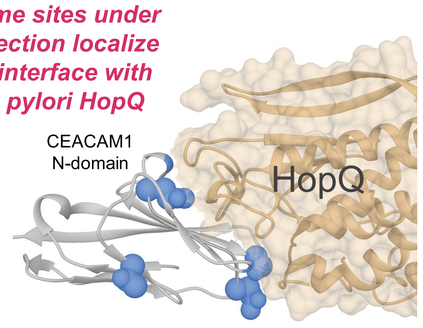
*Sites under selection are concentrated in the N-domains**

domain targeted by bacterial adhesins



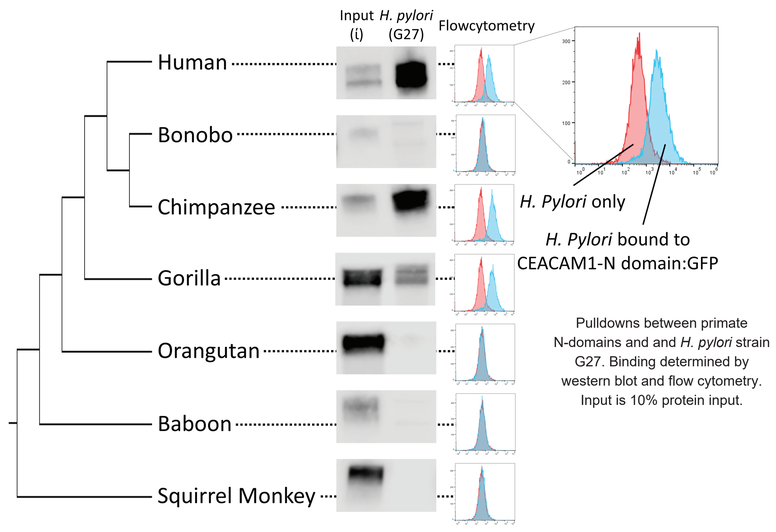
*** based on NS-sites analysis in PAML**

Some sites under selection localize to interface with H. pylori HopQ



Highlighted sites indicated as under selection by phylogenetic analysis programs PAML, FEL[6] and MEME[7]

Sequence differences between primate N-domain determines binding with H. pylori HopQ



Pulldowns between primate N-domains and and *H. pylori* strain G27. Binding determined by western blot and flow cytometry. Input is 10% protein input.

Bonobo CEACAM1 N-domain is unable to bind HopQ and has unusually divergend protein sequence

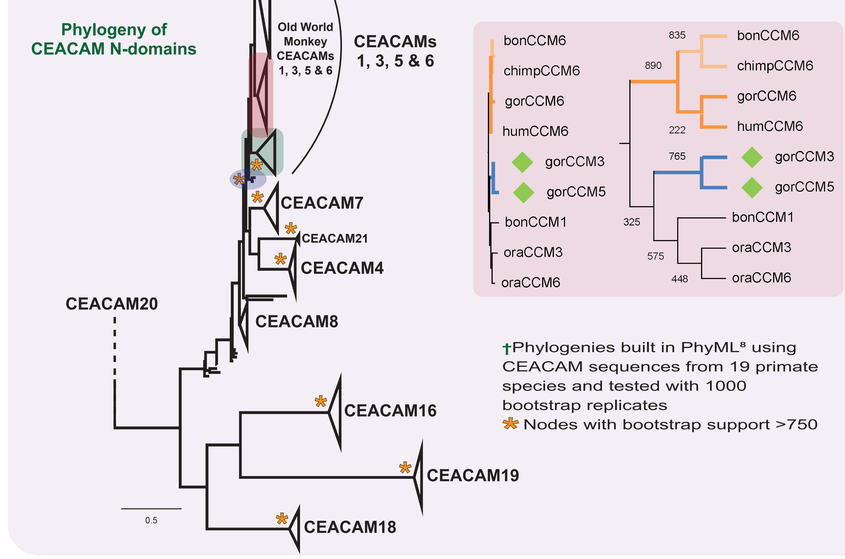
```
Bonobo  KLTIESTPFNVAEGKEVLLLTHNLPQNHIGYIWYKGERVDGNRIIVAHAIQNQQTTRGPAHSGRETVYPN
Human   QLTTESMPFNVAEGKEVLLLVHNLPQQLFGYSWYKGERVDGNRQIVGYAIGTQQATPGPANSGRETIYPN
Chimp   QLTTESMPFNVAEGKEVLLLVYNLPQQLFGYSWYKGERVDGNRQIVGYVIGTQQATPGPAYSGRETTYPN
```

Rapid accumulation of amino acid changes in bonobo CEACAM1 could be due to gene conversion within the CEACAM N-domain

In phylogeny of full length sequences homologous CEACAM proteins generally group together with high bootstrap support

In phylogeny of only CEACAM N-domains...

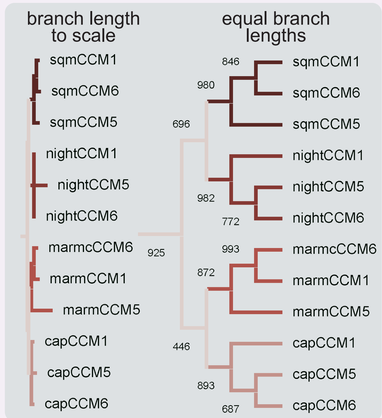
- slower evolving CEACAMs group with homologous sequences with high bootstrap support
- ***Several clades support recent gene conversion among paralogous N-domains***
- CEACAMs 1, 3, 5 & 6 cluster together
  - Often, not alway, with low support
- Origin of bonobo CEACAM1 still unclear
  - May be the result of multiple conversion events



Clades supporting gene conversion

†Phylogenies built in PhyML[8] using CEACAM sequences from 19 primate species and tested with 1000 bootstrap replicates

✱ Nodes with bootstrap support >750



By swapping similar, but divergent N-domains CEACAMs antagonized by pathogens may be able to evade exploitation while maintaining essential functions.

1. Correa P, Piazuelo MB. Evolutionary history of the Helicobacter pylori genome: Implications for gastric carcinogenesis. Gut Liver. 2012;6: 21–28. doi:10.5009/gnl.2012.6.1.21
2. Ishaq S, Nunn L. Helicobacter pylori and gastric cancer: A state of the art review. Gastroenterol Hepatol from Bed to Bench. 2015;8: S6–S14.
3. Javaheri A, Kruse T, Moonens K, Mejías-luque R, Debraekeleer A, Asche CI, et al. Helicobacter pylori adhesin HopQ engages in a virulence-enhancing interaction with human CEACAMs. Nat Microbiol. 2016;17: 16189. doi:10.1038/nmicrobiol.2016.189
4. Gray-Owen SD, Blumberg RS. CEACAM1: Contact-dependent control of immunity. Nat Rev Immunol. 2006;6: 433–446. doi:10.1038/nri1864
5. Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586–1591.
6. Kosakovsky Pond SL, Frost SD. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular biology and evolution. 2005 May 1;22(5):1208-22.
7. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SL. Detecting individual sites subject to episodic diversifying selection. PLoS genetics. 2012 Jul;8(7).
8. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology. 2010 May 1;59(3):307-21.

Learn more at: barberlab.org

----------Contact----------

Baker ECP: ebaker6@uoregon.edu

Barber MF: mfbarber@uoregon.edu

Work in the Barber laboratory has been supported by NIH grants R00GM115822 and R35GM133652 as well as a Faculty Scholar Award from the Donald E. and Delia B. Baxter Foundation.