

Gene cooption without duplication during the evolution of a male-pregnancy gene in pipefish

April Harlin-Cognato, Eric A. Hoffman, and Adam G. Jones

PNAS published online Dec 8, 2006;
doi:10.1073/pnas.0603000103

This information is current as of December 2006.

Supplementary Material

Supplementary material can be found at:
www.pnas.org/cgi/content/full/0603000103/DC1

This article has been cited by other articles:
www.pnas.org#otherarticles

E-mail Alerts

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Rights & Permissions

To reproduce this article in part (figures, tables) or in entirety, see:
www.pnas.org/misc/rightperm.shtml

Reprints

To order reprints, see:
www.pnas.org/misc/reprints.shtml

Notes:

Gene cooption without duplication during the evolution of a male-pregnancy gene in pipefish

April Harlin-Cognato^{†‡}, Eric A. Hoffman[§], and Adam G. Jones[†]

[†]Department of Biology, Texas A&M University, 3258 TAMU, College Station, TX 77843-3258; and [§]Department of Biology, University of Central Florida, 4000 Central Florida Boulevard, Orlando, FL 32816

Edited by Morris Goodman, Wayne State University School of Medicine, Detroit, MI, and approved October 5, 2006 (received for review April 14, 2006)

Comparative studies of developmental processes suggest that novel traits usually evolve through the cooption of preexisting genes and proteins, mainly via gene duplication and functional specialization of paralogs. However, an alternative hypothesis is that novel protein function can evolve without gene duplication, through changes in the spatiotemporal patterns of gene expression (e.g., via cis-regulatory elements), or functional modifications (e.g., addition of functional domains) of the proteins they encode, or both. Here we present an astacin metalloprotease, dubbed *patriscin*, which has been coopted without duplication, via alteration in the expression of a preexisting gene from the kidney and liver of bony fishes, for a novel role in the brood pouch of pregnant male pipefish. We examined the molecular evolution of *patriscin* and found conservation of astacin-specific motifs but also several positively selected amino acids that may represent functional modifications for male pregnancy. Overall, our results pinpoint a clear case in which gene cooption occurred without gene duplication during the genesis of an evolutionarily significant novel structure, the male brood pouch. These findings contribute to a growing understanding of morphological innovation, a critically important but poorly understood process in evolutionary biology.

novel trait evolution | *patriscin* | Syngnathidae

Evolutionary innovation has been defined as “the origin of a novel body part which may serve a novel function or specialize in a function that was already performed in the ancestral lineage but without a dedicated organ” (ref. 1, p. 581). In seahorses and pipefishes (family Syngnathidae) males carry their embryos on their ventral surface, either exposed to the environment or enclosed in a fleshy brood pouch (Fig. 1). The brood pouch is a clear example of an evolutionary innovation: syngnathids are the only lineage to have evolved a morphological structure that allows males to become pregnant. In many species of syngnathid fishes, the brood pouch is a complex organ composed of highly vascularized epithelial tissue that forms a honeycomb matrix encapsulating individual eggs during gestation (Fig. 1) (2). This placenta-like tissue apparently supplies nutrients to developing embryos (3) in a manner analogous to the placenta of female mammals. The pouch is lined with cells rich in mitochondria (CRMs) (Fig. 1) (2, 4) that transfer ions between the brood pouch fluid and the male bloodstream, maintaining an osmotically neutral environment during the early stages of gestation (4–6). A phylogeny of the Syngnathidae suggests two independent origins of the brood pouch based on its location on either the abdomen (Gastrophori) or the tail (Urophori) of the male (7), and within these two lineages there is considerable variation in the degree to which the pouch encloses developing embryos, the complexity of the “pseudo-placenta,” and the structure of brood pouch folds (7).

Clearly, the evolution of a structure as complex as a brood pouch required the coevolution of a multitude of genes, but little is known about the genomic changes associated with the evolution of the developmental and physiological processes that allow such a unique feature to function. The conservation of proteins and motifs across the tree of life suggests that the appropriation, or cooption, of preexisting genes and proteins plays an important

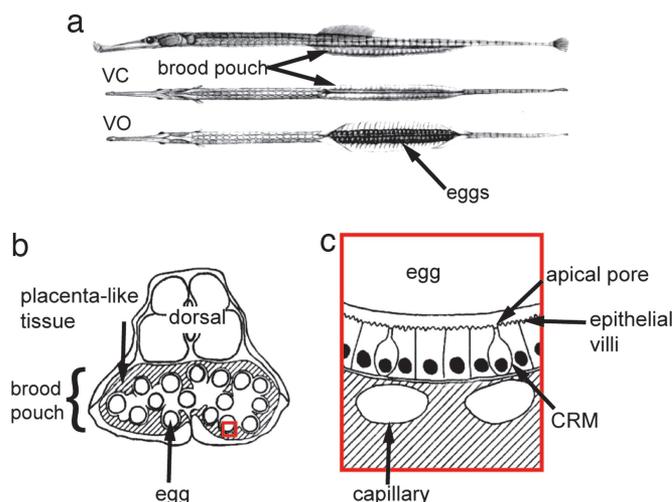


Fig. 1. Diagram of brood pouch morphology of tail-brooding pipefish. (a) Three views of the brood pouch. VC, ventral, closed; VO, ventral, opened. (b) A cross-section of a gravid brood pouch, showing the placenta-like epithelium with capsules for individual embryos. Note the complete closure of the pouch folds. (c) A closer view shows details of the structure of the pseudo placenta and the egg-placenta interface. The egg lies on the surface of villous epithelial cells interspersed with CRMs that have an apical pore opening into the lumen and a base that lies close to capillaries. These CRMs are thought to be the conduits by which ions are transported out of the brood pouch to maintain an environment suitable for developing embryos. This figure is adapted from Kornienko (41).

role in the evolution of novel traits (8). Gene cooption occurs when a new function is derived for an old gene, usually through changes in expression patterns and/or functional modifications of the protein it encodes (8). Because protein evolution is necessarily constrained to retain ancestral function, it has long been considered that duplications of genes, gene segments, or entire genomes are the primary source of material for the cooption of proteins during the evolution of novel traits (9, 10). However, it is not clear how duplicated genes survive the accumulation of point mutations while making the transition to a new function (10) or how these new functions are acquired (11). A number of studies have shown that duplicated genes can

Author contributions: A.H.-C., E.A.H., and A.G.J. designed research; A.H.-C. and E.A.H. performed research; A.H.-C. analyzed data; and A.H.-C. and A.G.J. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS direct submission.

Abbreviations: CRMs, cells rich in mitochondria; BF, Bayes factor; REL, random effects likelihood.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EF060269–EF060286).

[†]To whom correspondence should be sent at the present address: Department of Zoology, Michigan State University, 203 Natural Sciences Building, East Lansing, MI 48824. E-mail: cognatoa@msu.edu.

© 2006 by The National Academy of Sciences of the USA

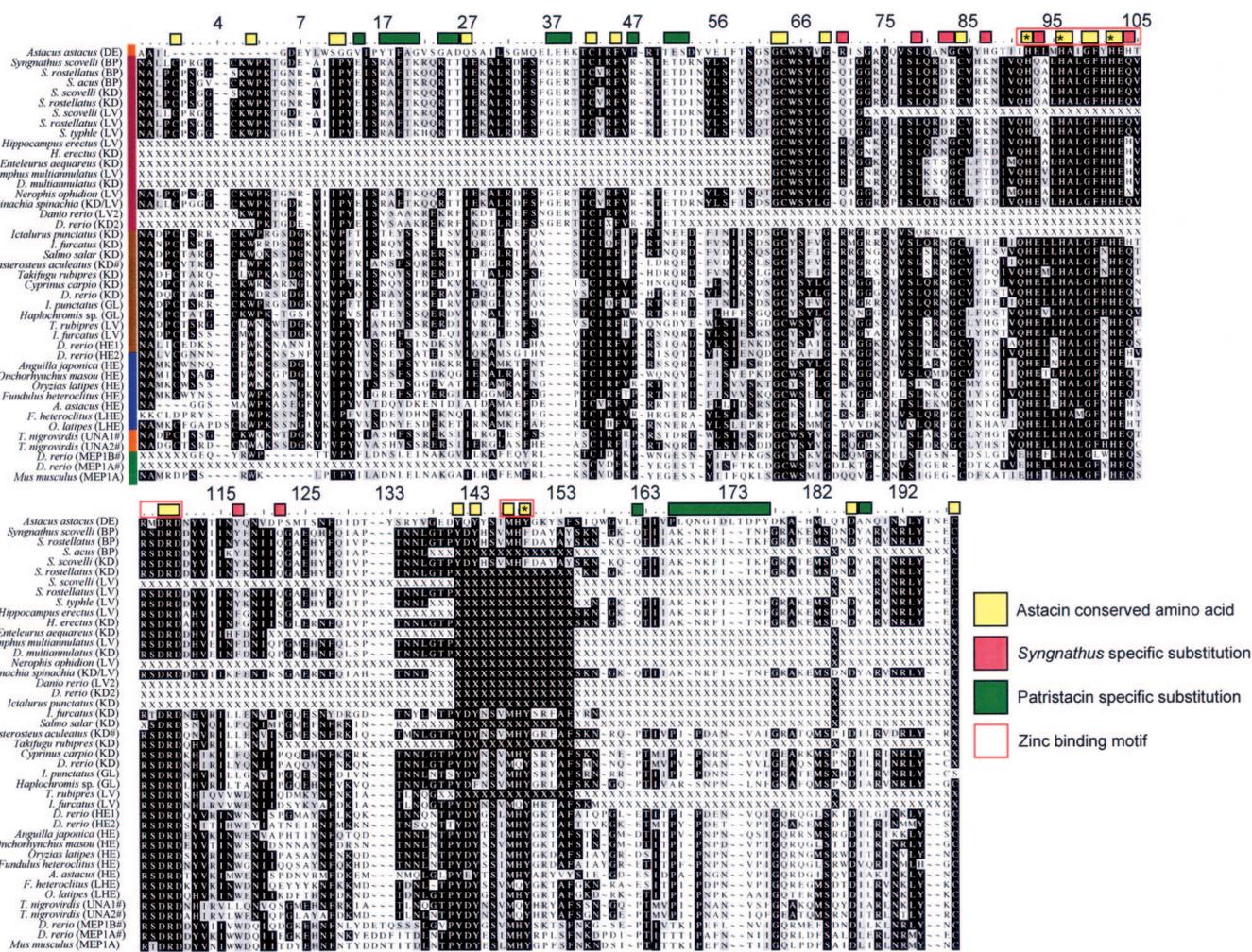


Fig. 2. Amino acid alignment of astacin mature peptides. Amino acids are numbered according to alignment with astacin, the type-protein of the subfamily. Shaded blocks designate amino acid positions that share $\geq 35\%$ similarity across all sequences. Astacins are characterized by four zinc bonds (indicated by asterisks) within two conserved zinc binding motifs (HEXXHXAGFXGEXRXDR, MHY) (17, 42). Colored bars to the left of the sequences identify astacin protein type as in Fig. 3. DE, digestive enzyme; BP, brood pouch; LV, liver; KD, kidney; GL, gill; HE, high hatching enzyme; LHE, low hatching enzyme; UNA, unknown tissue type; MEP, meprin subunits α (A) and β (B); #, proteins whose tissue-specific association was inferred from phylogenetic analysis; X, missing data; ~, gap inserted for alignment.

survive in the genome by acquiring a new function (neofunctionalization) or by partitioning their function among gene copies (subfunctionalization) (10, 12, 13). Some examples of these processes include the addition of functional domains to an existing protein, or changes in cis-regulatory elements that alter the spatiotemporal patterns of protein expression (13). However, only a fraction of gene duplications are thought to make the transition to new functions (14), so it is unlikely that biological diversity can be explained entirely by the processes of duplicate gene evolution.

As an alternative to the strict gene duplication model, gene cooption may occur through the alteration of spatial patterns of expression of existing genes. Under this model, functional promiscuity and conformational diversity of proteins play an important role in the evolution of novelty (11, 15). Proteins are evolutionarily labile, and variation in protein conformation and substrate affinity in different physicochemical environments and in the presence of a variety of substrates allows a preexisting protein to rapidly acquire a functional role associated with a novel structure (11, 15, 16), without gene duplication or changes to the original protein function.

Our goals in the present study were to identify genes involved in male pregnancy and to investigate the molecular evolution of

one of these male-pregnancy genes to distinguish between different models of gene cooption. In particular, we wanted to test the hypotheses that gene cooption during the evolution of male pregnancy was accompanied by gene duplication, and that genes involved in male pregnancy experienced a history of positive selection associated with their recent cooption to the brood pouch. Under the duplication model, we expect to recover a brood-pouch-specific gene lineage immediately after a duplication event, whereas under the nonduplication model we expect there to be no indication of tissue-specific gene evolution. Positive selection is possible under both duplication and nonduplication models and would indicate potential modifications to the protein associated with male pregnancy. However, the interpretation of these results would differ under each model.

Results and Discussion

To investigate the genetic basis of novel gene function in the brood pouch of pipefishes and seahorses, we used subtractive suppression hybridization of cDNA libraries from the brood pouch tissue of pregnant and nonpregnant male Gulf pipefish (*Syngnathus scovelli*). This approach allowed us to construct a cDNA library enriched for genes that were up-regulated in the

significance determined from critical values of the χ^2 distribution with estimated degrees of freedom as described (32). If the model with positive selection had a significantly better fit to the data, the posterior probability for positive selection at each site was estimated (33). A posterior probability $\geq 90\%$ was considered moderate support and a posterior probability $\geq 95\%$ was considered strong support that a site is under positive selection. In addition, we tested for positive selection among sites with the random effects likelihood (REL) method (34) via the Datamonkey (35) adaptive evolution server (www.datamonkey.org). The REL method is a modified parsimony-based method, which allows for rate heterogeneity in both dN and dS (36). REL analysis calculates two Bayes factors (BF), one for negative selection (dN < dS) and one for positive selection (dN > dS). We considered a BF ≥ 50 as moderate evidence and a BF ≥ 100 as strong evidence that a site is under positive selection. Whether maximum likelihood or parsimony-based methods are better for detecting sites under positive selection is a subject of debate (37, 38). To circumvent this controversy, we usually considered sites to be under positive selection only if (i) they met the initial BF and posterior probability limits and (ii) were identified in both REL and PAML analyses. Ten of the sites identified with REL ($n = 12$) and PAML ($n = 14$) methods met these criteria. We also identified two sites from the REL analysis

that exhibited very strong evidence for positive selection (BF $\gg 100$) and hence included them in the final count of 12 positively selected sites.

Protein 3D Model. We modeled the 3D structure of patristacin via the Swiss-Model homology modeling server (39) (<http://swissmodel.expasy.org>) with the fully resolved crystal structure of the *Astacus astacus* digestive enzyme (astacin) as a template. Amino acid substitutions in patristacin were compared with aligned substrate-binding and conserved structural residues in astacin. These sites and those under positive selection were visualized and annotated in Cn3D v. 4.1 (40).

Molecular Kits. TOPO TA (cloning), GeneRacer (RACE), SuperScript III (reverse transcription), and TRIzol (RNA isolation) were from Invitrogen (Carlsbad, CA). PCR-Select (subtraction suppression hybridization) was from Clontech (Mountain View, CA).

We acknowledge K. B. Mobley, C. Partridge, Z. Cress, and C. Small for sample collection (pipefishes and sticklebacks). S. Coleman, R. Honeycutt, T. Ward, K. B. Mobley, and two anonymous reviewers provided helpful comments on previous drafts. Funding was provided by a grant from the National Science Foundation (to A.G.J.).

1. Wagner GP, Lynch VJ (2005) *J Exp Zool B* 304:580–592.
2. Carcupino M, Baldacci A, Mazzini M, Franzoi P (2002) *J Fish Biol* 61:1465–1480.
3. Haresign TW, Shumway SE (1981) *Comp Biochem Physiol A Physiol* 69:603–604.
4. Watanabe S, Kaneko T, Watanabe Y (1999) *Cell Tissue Res* 295:141–149.
5. Quast WD, Howe NR (1980) *Comp Biochem Physiol A Physiol* 67:675–678.
6. Azarello MY (1991) *Bull Mar Sci* 49:741–747.
7. Wilson AB, Vincent A, Ahnesjo I, Meyer A (2001) *J Hered* 92:159–166.
8. True JR, Carroll SB (2002) *Annu Rev Cell Dev Biol* 18:53–80.
9. Ohno S (1970) *Evolution by Gene Duplication* (Springer, Berlin).
10. Lynch M, Katju V (2004) *Trends Genet* 20:544–549.
11. James LC, Tawfik DS (2003) *Trends Biochem Sci* 28:361–368.
12. de Souza FSJ, Bumashny VF, Low MJ, Rubinstein M (2005) *Mol Biol Evol* 22:2417–2427.
13. Force A, Lynch M, Pickett FB, Amores A, Yan Y-L, Postlethwait JH (1999) *Genetics* 151:1531–1545.
14. Lynch M, Conery JS (2000) *Science* 290:1151–1155.
15. Jensen RA (1976) *Annu Rev Microbiol* 30:409–425.
16. Copley SD (2003) *Curr Opin Chem Biol* 7:265–272.
17. Bond JS, Benyon RJ (1995) *Protein Sci* 4:1247–1261.
18. Titani K, Torff JJ, Hormel S, Kumar S, Walsh KA, Rodi J, Neurath H, Zwilling R (1987) *Biochemistry* 26:222–226.
19. Inohaya K, Yasumasu S, Ishimaru M, Ohyama A, Iuchi I, Yamagami K (1995) *Dev Biol* 171:374–385.
20. Yasumasu S, Iuchi I, Yamagami K (1988) *Zool Sci* 5:191–195.
21. Tabas JA, Zasloff M, Wasmuth JJ, Emanuel BS, Altherr MR, McPherson JD, Wonzney JM, Kaplan FS (1991) *Genomics* 9:283–289.
22. Ishmael FT, Norcum MT, Benkovic SJ, Bond JS (2001) *J Biol Chem* 276:23207–23211.
23. Kijimoto T, Watanabe M, Fujimura K, Nakazawa M, Murakami Y, Kuratani S, Kohara Y, Gojobori T, Okada N (2005) *Mol Biol Evol* 22:1649–1660.
24. Hung C-H, Huang H-R, Huang C-J, Huang F-L, Chang G-D (1997) *J Biol Chem* 272:13772–13778.
25. Bordo D, Argos P (1991) *J Mol Biol* 217:721–729.
26. Foskett JK, Scheffey C (1982) *Science* 215:164–166.
27. Varsamos S, Nebel C, Charmanier G (2005) *Comp Biochem Physiol A Physiol* 141:401–429.
28. Diatchenko L, Lau Y-FC, Campbell AP, Chenchik A, Moqadam F, Huang B, Lukyanov S, Lukyanov K, Gurskaya N, Sverdlov ED, Siebert PD (1996) *Proc Natl Acad Sci USA* 93:6025–6030.
29. von Stein OD, Thies WG, Hofmann M (1997) *Nucleic Acids Res* 25:2598–2602.
30. Poirot O, Suhre K, Abergel C, O'Toole E, Notredame C (2004) *Nucleic Acids Res* 32:W37–W40.
31. Swofford DL (1999) *PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).
32. Nielsen R, Yang ZH (1998) *Genetics* 148:929–936.
33. Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) *Genetics* 155:431–449.
34. Kosakovsky Pond SL, Frost SDW (2005) *Mol Biol Evol* 22:1208–1222.
35. Kosakovsky Pond SL, Frost SDW (2005) *Bioinformatics* 21:2531–2533.
36. Kosakovsky Pond SL, Muse SV (2005) *Mol Biol Evol* 22:2375–2385.
37. Suzuki Y, Nei M (2001) *Mol Biol Evol* 18:2179–2185.
38. Anisimova M, Bielawski JP, Yang Z (2002) *Mol Biol Evol* 19:950–958.
39. Schwede T, Kopp J, Guex N, Peitsch M (2003) *Nucleic Acids Res* 31:3381–3385.
40. Hogue C (1997) *Trends Biochem Sci* 22:314–316.
41. Kornienko ES (2001) *Russ J Mar Biol* 27:S15–S26.
42. Rawlings ND, Barrett AJ (1995) *Methods Enzymol* 248:183–228.