Available online at www.sciencedirect.com
ScienceDirect
Gene 392 (2007) 47–58
GENE
SECTION
EVOLUTIONARY GENOMICS
www.elsevier.com/locate/gene

A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs

Todd A. Castoe [a], Tricia Stephens [a], Brice P. Noonan [b,1], Cristina Calestani [a,*]

[a] *Department of Biology, University of Central Florida, 4000 Central Florida Blvd., Orlando, FL 32816-2368, USA*
[b] *Department of Integrative Biology, Brigham Young University, 401 WIDB, Provo, UT 84604, USA*

Received 30 June 2006; received in revised form 3 November 2006; accepted 10 November 2006
Available online 17 November 2006
Received by A. Bernardi

Abstract

Type I polyketide synthases (PKSs), and related fatty acid synthases (FASs), represent a large group of proteins encoded by a diverse gene family that occurs in eubacteria and eukaryotes (mainly in fungi). Collectively, enzymes encoded by this gene family produce a wide array of polyketide compounds that encompass a broad spectrum of biological activity including antibiotic, antitumor, antifungal, immunosuppressive, and predator defense functional roles. We employed a phylogenomics approach to estimate relationships among members of this gene family from eubacterial and eukaryotic genomes. Our results suggest that some animal genomes (sea urchins, birds, and fish) possess a previously unidentified group of *pks* genes, in addition to possessing *fas* genes used in fatty acid metabolism. These *pks* genes in the chicken, fish, and sea urchin genomes do not appear to be closely related to any other animal or fungal genes, and instead are closely related to *pks* genes from the slime mold *Dictyostelium* and eubacteria. Continued accumulation of genome sequence data from diverse animal lineages is required to clarify whether the presence of these (non-*fas*) *pks* genes in animal genomes owes their origins to horizontal gene transfer (from eubacterial or *Dictostelium* genomes) or to more conventional patterns of vertical inheritance coupled with massive gene loss in several animal lineages. Additionally, results of our broad-scale phylogenetic analyses bolster the support for previous hypotheses of horizontal gene transfer of *pks* genes from bacterial to fungal and protozoan lineages.

Keywords: Fatty acid synthases; Gene loss; Horizontal gene transfer; Phylogeny; Sea urchin; *Strongylocentrotus purpuratus*; *SpPks*

1. Introduction

Polyketide synthases (PKSs) are multifunctional enzymes involved in the biosynthesis of a wide range of polyketide compounds (Hopwood, 1997; Hopwood and Sherman, 1990; Staunton and Weissman, 2001). Many diverse polyketides are synthesized by bacteria and fungi and have antibiotic or mycotoxic properties (e.g., erythromycin, rifamycin and actinorhodin). Polyketide compounds are also synthezised by plants, in which they have diverse functions including roles in flower pigmentation, pathogen defense (phytoalexins), UV and visible light exposure response, and symbiotic plant–pathogen interactions (Schroder et al., 1998; Winkel-Shirley, 2002). The diversity of biological activity of polyketides has made these secondary metabolites, and the PKS proteins that synthesize them, an important focus of biopharmaceutical research.

PKSs and fatty acid synthases (FASs) have similar protein domains and their biosynthetic pathways have several features in common. PKS/FASs typically catalyze multiple successive rounds of condensation of simple carbon units (usually acetyl-CoA and malonyl-CoA) to build a β-keto chain. In fatty acid biosynthesis, however, there is a complete reduction of the keto

Abbreviations: AMP, AMP binding domain; AT, acyl transferase domain; BMCMC, Bayesian Markov chain Monte Carlo phylogenetic analysis; BSS, bootstrap support; CHS, chalcone synthase domain; DH, dehydratase domain; ER, enoyl reductase domain; EST, expressed sequence tag; *fas*, fatty acid synthase gene; FAS, fatty acid synthase protein; HGT, horizontal gene transfer; KR, ketoreductase domain; KS, ketoacyl synthase domain; ME, methylase domain; MP, maximum parsimony; *pks*, polyketide synthase gene; PKS, polyketide synthase protein; PP, phosphopantetheine attachment site; PPS, posterior probability support; TE, thioesterase domain.

* Corresponding author. Tel.: +1 407 823 4504; fax: +1 407 823 5769.
E-mail address: calestac@mail.ucf.edu (C. Calestani).
[1] Current address: Department of Biology, Duke University, Box 90338, Durham, NC 27708, USA.

0378-1119/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2006.11.005

groups with the production of completely saturated carbon chains, while these remain unreduced or partially reduced in polyketides.

PKSs have been classified as types I, II, and III, and our study focuses primarily on proteins within type I (for a review of PKS classification, see Moss et al., 2004). Type I PKSs contain, within a multifunctional polypeptide, all the enzymatic activities necessary for one cycle of β-keto chain elongation and processing, and may be either modular (mostly in bacteria) or iterative (in fungi). In modular PKSs, each polypeptide includes one or multiple modules, and each module is responsible for one round of condensation and β-keto chain processing. Each catalytic domain of modular type I PKSs is used only once during the biosynthetic process. In contrast, iterative type I PKSs are monomodular and conduct multiple rounds of chain elongation and β-keto chain processing using their catalytic domains multiple times (for illustrations of polyketide synthesis in relation to protein domain structure see: Hopwood and Sherman, 1990; Hopwood, 1997; Moss et al., 2004; Jenke-Kodama et al., 2005). Type I PKSs and FASs often share a conserved structure that includes the following functional domains: a ketoacyl synthase (KS), acyl transferase (AT), and phosphopantetheine attachment site (PP; also known as an acyl carrier protein domain; Hopwood, 1997). FASs contain additional domains including ketoreductase (KR), dehydratase (DH), enoyl reductase (ER), and thioesterase (TE) domains; these domains may or may not be present in PKSs.

In contrast to type I PKSs, the enzymatic activities for the β-keto chain elongation and processing in type II PKSs are present in separate polypeptides, and each domain is used iteratively. Type III PKSs, or calchone synthases, were originally identified in plants and have been recently been isolated also from several bacteria (Moore and Hopke, 2001; Gross et al., 2006). In contrast to type I and type II polyketide biosynthesis, the β-keto chain is elongated and processed at a single multifunctional active site in type III PKSs, and type III proteins do not possess a PP domain.

Recent phylogenetic evidence (Kroken et al., 2003; Jenke-Kodama et al., 2005) has suggested that animal FASs are evolutionarily nested within the type I *pks* gene family, implying that *fas* genes represent an evolutionary subset of type I *pks* genes that occur in animals. Polyketide compounds, used as defensive mechanisms to deter predation, have been isolated from some marine invertebrates (e.g., sponges and mollusks; Garson, 1989), although these polyketides were subsequently found to be produced by bacterial symbionts (and not encoded in the animal genomes). Aside from *fas* genes, no other forms of *pks* genes are currently known to occur in the genomes of animals except for an enigmatic *pks* identified in sea urchins (Calestani et al., 2003).

Calestani et al. (2003) isolated a *pks* gene (*SpPks*) expressed in pigment cells of sea urchin (*Strongylocentrotus purpuratus*) embryos. Sea urchin larval pigment cells are a subset of mesodermal cells with mesenchymal properties. In their final position, pigment cells are embedded in the ectoderm of the larva. The pigment produced by pigment cells is a naphthoquinone called echinochrome that has characteristics typical of a polyketide compound (McLendon, 1912; Kuhn and Wallenfells, 1940; Griffiths, 1965). *SpPks* was found to be required for the biosynthesis of the pigment echinochrome (Calestani et al., 2003), and sea urchin embryos lacking SpPKS (knock-down) develop pigment cells, but they appear unpigmented (albino phenotype; Calestani et al., 2003). The sequence of isolated cDNA clones of *SpPks* demonstrated that this gene encodes a multifunctional polypeptide containing the following series of conserved domains: KS–AT–DH–ER–PP (Calestani et al., 2003). The presence of these domains indicates that SpPKS belongs to the type I class of PKSs. In accordance with these data, a biochemical study by Salaque et al. (1967) showed that acetic acid molecules are utilized as precursors in the biosynthesis of echinochrome A in sea urchin. Currently, the function of sea urchin echinochrome is not completely understood, but may include roles in immuno-defense (Service and Wardlaw, 1984).

Initial comparisons of SpPKS with other known proteins (via Blast searches of the NCBI database) revealed particularly low similarity with *Caenorhabditis*, *Drosophila*, and human FASs, and there appeared to be no obvious orthologs of *SpPks* in other animal genomes. Furthermore, SpPKS appeared to share high similarity with unrelated fungal and bacterial PKSs. The enigmatic similarities of SpPKS with non-animal PKSs, based on initial analyses, lead us to ask specific questions about the evolutionary origins of this gene, including: (*i*) Is there phylogenetic evidence for orthologs of *SpPks* in other animal genomes available? (*ii*) Is *SpPks* more closely related to animal *fas* genes, the most similar animal genes to *SpPks* based on initial BLAST results? (*iii*) Are the closest relatives of *SpPks* non-animal genes in fungi or bacteria? (*iv*) Is gene loss, or horizontal gene transfer (HGT), a plausible explanation to account for the evolutionary origins of *SpPks*?

In order to address these questions we assembled a large collection of amino acid sequences of type I PKSs and FASs from eubacteria and eukaryotes that shared similarity with SpPKS, and conducted phylogenetic analyses to estimate relationships among these proteins. In addition to our specific questions regarding the evolutionary history and relations of *SpPks*, the taxonomic breadth of our analysis also provided an optimal opportunity to address more general questions about the evolution of eubacterial and eukaryotic *pks* and *fas* genes. We exploit our broad-scale phylogenetic estimates to address the following additional questions: (*i*) What are the relationships among major groups of eubacterial and eukaryotic type I PKS and FAS proteins? (*ii*) Across the phylogeny of PKS/FASs, is there any evidence for HGT and/or differential gene loss in eukaryotic lineages?

2. Materials and methods

2.1. Annotation of sea urchin pks genes and identification of conserved protein domains

The predicted exon–intron structure and protein sequence of *SpPks* was verified by tiling expressed sequence tag (EST) data (partial cDNA clones) over the corresponding genome sequence cloned into a bacterial artificial chromosome. This gene has a total of seven exons (2424 predicted codons), two of which are

quite large (exon two is ~1.5 kb, and exon four is ~5 kb). We have reannotated the initial prediction of this gene based on our EST data (NCBI accession XM_788471). A very similar *pks* (*LvPks*) from another species of sea urchin, *Lytechinus variegatus*, was also identified based on alignment of *SpPks* exons with this species' EST sequences and corresponding genomic sequence cloned into a bacterial artificial chromosome. This gene prediction (*LvPks*), with accompanying annotation, was submitted to the NCBI database (accession AC131501).

Conserved domains of proteins were identified using the Protein Families Database of Alignments and HMMs (PFAM) tools available on the Sanger Institute server (http://www.sanger.ac.uk/Software/Pfam/). Here, we use the following abbreviations for protein domains of PKSs: KS — ketoacyl synthase, AT — acyl transferase, PP — phosphopantetheine attachment site (also known as acyl carrier protein domain), DH — dehydratase, KR — ketoreductase, ER — enoyl reductase, CHS — chalcone synthase, AMP — AMP binding, ME — methylase and TE — thioesterase.

2.2. Data retrieval from genomic databases

We searched for PKS and FAS proteins that were similar to SpPKS using BLASTP, BLASTN, and TBLASTN searches against complete and incomplete genomes of eubacteria and eukaryotes deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/). For BLAST searches, we used one of the following four sequences: the complete amino acid or nucleic acid sequence of SpPKS, the amino acid sequence of only the KS domain of SpPKS, and the amino acid sequence of human FAS (NCBI accession NP_004095).

We thoroughly searched each group of organisms (e.g., plants, fungi, eubacteria, invertebrate animals, etc.) for sequences similar to SpPKS by conducting individual BLAST searches against each organismal group independently. Several organisms of interest (e.g., *Xenopus tropicalis*; see Results Section 3.1) with draft genomes or EST sequences available were not accessible on NCBI, and BLASTP and TBLASTN searches were conducted independently on these genomes via their respective web servers (links available from the NCBI website at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=genomeprj). We also searched the incomplete invertebrate genomes or EST databases of: the sea urchin *L. variegatus*, the urochordate *Ciona intestinalis*, the hemichordate *Saccoglossus kowalevskii*, and the cnidarians *Hydra magnipapillata* and *Nematostella vectensis*.

In addition to the collection of PKS sequences obtained from searching genomic databases, we incorporated all amino acid sequences included in a recent study on PKSs in ascomycete fungi (Kroken et al., 2003; and removed redundant sampling from our BLAST results). Outgroup sequences were obtained from bacterial type II and bacterial and mitochondrial ketoacyl-ACP-synthetases, as in Kroken et al. (2003).

2.3. Amino acid sequence alignment

Amino acid alignment was accomplished in an iterative fashion with rounds of automated alignment in CLUSTALW (using the BOLSUM62 matrix) followed by manual readjustment, and partial automated realignment of selected sequences or regions. A core motif among PKS proteins includes the typically adjacent KS and AT domains, and these domains represented the most conserved regions among PKSs. Accordingly, our final alignment for phylogenetic analyses included only these two regions (KS and AT domains). Sites in alignments where homology was ambiguous were conservatively excluded prior to phylogenetic analyses; a majority of excluded sites represented linker regions between the KS and AT domains and distal/proximal portions of the KS and AT domains. This final alignment included a total of 178 PKS protein sequences (5 of which were outgroups), aligned over 552 amino acid positions. All protein sequences used for phylogenetic analyses are provided, with NCBI accession numbers in the Online Supplementary Table (S1).

We expected the combined KS+AT domain alignment to produce superior estimates of phylogeny given that this combination of domains provides more phylogenetic characters. There is, however, some evidence that KS and AT domains may not share a common evolutionary history in some eubacteria (Jenke-Kodama et al., 2005). If this situation was common among *pks* genes, this may obscure estimates of the gene family phylogeny if such different phylogenetic signal (from the two domains) were mixed to estimate a single tree. To address this concern, we also conducted phylogenetic analyses based on a reduced alignment that only included 381 amino acids solely from the KS domain (the larger and more conserved of the two domains). Phylogeny estimates based on this smaller KS domain alignment were compared to the larger KS+AT domain analyses to verify the consistency among phylogeny estimates.

2.4. Phylogenetic analyses

We used two different methods to reconstruct the phylogeny of type I PKS and FAS proteins, maximum parsimony (MP) and Bayesian Markov chain Monte Carlo phylogenetic analyses (BMCMC). Using both methods provides a convenient way to cross-validate phylogenetic estimates since these two methods have inherently different strengths and intrinsic assumptions about the evolutionary process. In general, however, we expect BMCMC analyses to produce superior estimates of phylogeny because it incorporates probabilistic models of amino acid substitution that should be less likely to be misled by the complexities of the amino acid substitution process (e.g., Huelsenbeck, 1995; Huelsenbeck and Crandall, 1997). Agreeing with a large body of literature on modeling nucleotide substitution, recent evidence suggests that Bayesian phylogenetic methods are particularly robust to branch-length differences in protein datasets, especially when among-site rate heterogeneity is accommodated using a gamma distribution (Mar et al., 2005).

Maximum parsimony phylogenetic analyses of both alignments (KS+AT and KS-only) were conducted in PAUP*4.0b10 (Sinauer, Sunderland, MA), using heuristic searches with nodal support assessed via non-parametric bootstrapping (Felsenstein,

1985; bootstrap support = BSS). Heuristic MP searches were conducted with all characters equally weighted, and with 200 random-taxon-addition sequences. Maximum parsimony bootstrap analyses were conducted with 200 bootstrap pseudo-replicates, with 10 random-taxon-addition sequence replicates per bootstrap pseudo-replicate. All phylogenies were rooted with multiple outgroup sequences: type II bacterial PKSs, and bacterial and mitochondrial ketoacyl-ACP-synthases (indicated in Supplemental Online Table S1).

Bayesian phylogenetic analyses were conducted in MRBAYES version 3.1 (Ronquist and Huelsenbeck, 2003) with nodal support assessed via bipartition posterior probability support (PPS), calculated by surveying the posterior distribution of tree estimates for the frequency of bipartition occurrence. The default settings were used for all priors and parameter estimates except where indicated below. In order to identify when BMCMC runs had reached stationarity, we monitored the trends in the cold chain likelihood scores and parameter values for evidence of reaching stable plateaus using Tracer (Rambaut and Drummond, 2003).

Two initial BMCMC analyses were conducted for each dataset on a single processor computer with 2×10^6 generations per run. For all BMCMC runs, we used mixed priors on amino acid substitution models with gamma-distributed rate variation (thus integrating model selection into the process of phylogenetic analysis), which resulted in 100% posterior probability for the Wagner model of amino acid substitution (Whelan and Goldman, 2001) as best-fitting both datasets. To assemble a large number (distribution) of post-burnin BMCMC estimates, we conducted multiple independent BMCMC runs and pooled the posterior (post-burnin) estimates from these, as this strategy has been shown to converge well on estimates derived from a smaller number of BMCMC runs conducted for a large number of successive generations (Castoe et al., 2004). Initial runs for both datasets indicated that BMCMC stationarity was reached by 1.2×10^6 generations. Final BMCMC analyses of the KS+AT domain dataset were conducted with forty independent BMCMC runs, each conducted for 2.5×10^6 generations (with the first 2×10^6 generations of each discarded as burnin). Final BMCMC analyses of the KS (only) domain dataset were conducted with six independent BMCMC runs, each conducted with 6×10^6 generations (with the first 2×10^6 generations of each discarded as burnin).

3. Results

3.1. Recovery of PKS proteins from genomic databases

We annotated and submitted to Genbank (see Supporting Online Table S1) all *pks* genes we identified that were previously either not annotated, or incompletely annotated in genomic databases. We found a total of three type I *pks*-like genes in the draft genome of the sea urchin *S. purpuratus*. The first, *SpPks* (XM_788471) has been reported previously (Calestani et al., 2003; although we have revised the annotation based on EST data), and the second, *SpPks2* (XM_796777), was previously predicted from the genomic sequence data, but otherwise unidentified. A third gene (*SpFas*, accession NW_790012) appeared to be an ortholog of animal *fas* genes based on protein domain structure, which was later confirmed by our phylogenetic analyses. We identified what appeared to be an ortholog of *SpPks* (*LvPks*; AC131501) from another sea urchin species *L. variegatus*, based on the unpublished sequence of a bacterial artificial chromosome clone. Other than *LvPks*, no additional *pks* or *fas* genes were found in the *L. variegatus* genomic data available.

We found no relevant BLAST results (bit scores < 50) in searches of plant or archeabacterial genomic databases. We recovered several hundred PKSs from eubacterial and fungal genomes, and selected 70 protein sequences with the highest BLAST scores (compared with SpPKS) from each group to include in our analyses. A (non-*fas*) *pks* gene has not been identified in humans, and our BLAST searches recovered only a single human *fas*, and no other potential *pks* orthologs in the human genome. In most other vertebrate genome databases we recovered only a single *fas*-like gene, as we did for *Caenorhabditis elegans* and *C. briggsae*. Like most vertebrate genomes, the draft genome assemblies of *Bos taurus*, *Felis catus*, *Pan troglodytes*, *Pongo pygmaeus*, and *Oryctolagus cuniculus* returned only a single *fas* ortholog that was very similar to other vertebrate FAS proteins, and these were not included in the final phylogenetic analyses. Insect genome databases (*Drosophila melanogaster* and *Anopheles gambiae*) returned three PKS-like predicted proteins, and it appears that all three represent alternative potential isoforms (i.e., alternative transcripts) from a single *fas*-like locus. We included all three of these predicted proteins in phylogenetic analyses because some of the predicted protein sequences were quite different. Orthologs of this insect *fas* locus were also found in the draft genomes of *Bombix mori* (the domestic silkworm) and *Apis mellifera* (the honey bee), but were not included in phylogenetic analyses because they were nearly identical to the other insect *fas*-like genes included (from *Drosophila* and *Anopheles*); no other *pks*-like genes were found in *Bombix* and *Apis*. We found a single FAS-like predicted protein in the frog, *X. tropicalis*, EST database, based on an incomplete EST. Preliminary phylogenetic analyses placed the *X. tropicalis* protein within animal FASs, but with a very long terminal branch (possibly indicating a poor quality EST or an incorrect translation). Due to these issues, this *Xenopus* putative FAS was omitted from final phylogenetic analyses. We identified two *pks*-like genes in the genomes of chicken (*Gallus gallus*) and fish (*Danio rerio* and *Tetraodon nigroviridis*). One of these genes from each genome appeared to be related to FASs (based on protein domain structure and high sequence similarity), whereas the second in each case appeared quite divergent from animal FASs. Our BLAST searches of several incomplete genomes, including the urochordate *C. intestinalis*, and the EST databases of the hemichordate *S. kowalevskii* (J. Gerhart, M. Kirschner, and C. Lowe, unpubl.) and of the cnidarians *H. magnipapillata* and *N. vectensis*, returned either no significant BLAST matches or estimated protein sequences (with very low BLAST scores) that were unalignable with other type I PKS protein sequences.

3.2. Results of phylogenetic analyses based on the KS+AT domain alignment

Maximum parsimony heuristic searches recovered 76 equally parsimonious trees of 25,022 steps in total length. Tree statistics for these trees are as follows: consistency index = 0.222, retention index = 0.480 and homoplasy index = 0.778. The strict consensus of the 76 optimal MP trees, along with nodal BSS values, is provided as a Supplementary Online Figure (Fig. S1).

The results of BMCMC phylogenetic analyses are summarized based on the combination of post-burnin estimates from forty independent MCMC runs (representing 20×10^6 post-burnin generations). The majority-rule consensus of these pooled post-burnin estimates is provided with PPS values of nodal support (Fig. 1). On this tree we have also indicated nominal clades or groups, and for each we have provided an indication of the typical (and alternative) domain structures of the proteins in each group (Fig. 1). We use the term "group" to indicate a collection of protein sequences that was not estimated to be a monophyletic clade based on either MP or BMCMC (e.g., the "Mixed Group" forms a clade based on the MP analyses, but does not form a clade based on the BMCMC analyses).

The two phylogenetic methods provided different estimates of basal relationships among major groups of PKS proteins included, although both phylogenetic estimates were associated with particularly low support values for these deep relationships. Other than the most basal (deep) phylogenetic relationships among major clades of type I PKS proteins, MP and BMCMC methods provided broadly similar estimates of relationships among protein sequences within these major clades (and for this reason, we do not show all the results of both analyses, but see Supporting Online Figure S1 for entire MP results). Based on available evidence that BMCMC analyses are expected to produce a superior estimate of phylogeny (as outlined above, Section 2.4), we generally favor the BMCMC phylogeny, but describe results derived from both methods.

A comparison of the major relationships recovered by MP and BMCMC analyses among nominal groupings is shown in Fig. 2a, b (following the nomenclature of Fig. 1). Bayesian phylogeny estimates suggest that the deepest divergence among type I PKSs exists between the Fungal Clade I and the remaining groups of PKSs, whereas MP estimated the deepest divergences to be among members of Bacterial Group II (Fig. 2a, b). The Animal FAS Clade was inferred to be the sister group to the Fungal Clade II by BMCMC, whereas this FAS clade formed the sister group of the Fungal Clade II and Fungal Clade based on MP. The Nested Fungal Clade was placed within Bacterial Group II by both methods, although the BMCMC estimate placed it more deeply nested among bacterial PKSs (Fig. 2a; see also Fig. S1A).

Both MP and BMCMC analyses resulted in topologies in which PKSs and FASs from animal genomes formed distantly related clades, separated by non-animal PKSs. Proteins from animal genomes which did not phylogenetically cluster within the Animal FAS Clade, along with other closely related sequences, are collectively referred to as the Mixed PKS Group. This Mixed PKS Group contained SpPKS, SpPKS2, and LvPKS, along with proteins from some vertebrates (birds and fish; Figs. 2b and S1B). The two phylogenetic methods implied slightly different relationships within the Mixed PKS Group (regarding the placement of several bacterial PKSs and SpPKS2). The relationships of these non-FAS animal PKSs (and related proteins) in the Mixed PKS Group are shown in Fig. 1B (based on BMCMC) and Fig. 2d (based on MP).

Although the two phylogenetic methods differed slightly in the estimated phylogenetic placement of these animal (non-FAS) PKS proteins (Figs. 1B and 2d), the differences between these estimates are not substantial. Both methods imply that the closest relative of the sea urchin proteins SpPKS (from *Strongylocentrotus*) and LvPKS (from *Lytechinus*) is a PKS from the slime mold *Dictyostelium* (PPS = 95; BSS < 50). In the BMCMC estimate (Fig. 1B), this group formed the sister clade (PPS = 64) to a cluster of PKSs from the chicken (*Gallus*) and fish (*Danio* and *Tetraodon*) genomes, and the sister group to this clade of animal and *Dictyostelium* PKSs was inferred to be a PKS from the marine plankton *Rhodopirellula* (Fig. 1B). In the MP estimate, the *Rhodopirellula* PKS formed the sister lineage to *Dictyostelium* and sea urchin PKSs, and the chicken and fish PKS clade formed the sister group to this clade (Fig. 2d). Surrounding this group in both estimates are several bacterial PKSs, which vary in phylogenetic placement slightly between BMCMC and MP estimates. Also, within this Mixed PKS Group, a second sea urchin PKS (SpPKS2, from *Strongylocentrotus*) is clustered variably between estimates with PKSs from the bacteria *Nocardia*, *Clostridium*, and *Microbulbifer* (in BMCMC; PPS < 50; Fig. 1B) or *Bordetella*, *Nitrosomonas*, and *Burkolderia* (in MP, BSS = 99; Fig. 2d).

In addition to the unexpected phylogenetic placement of (non-FAS) animal PKSs, both MP and BMCMC estimates agree in the placement of several fungal and protozoan PKSs nested within larger clades of bacterial PKSs (see black boxes in Figs. 1A and S1A). Both MP and BMCMC estimates placed a PKS from the fungus *Cochliobolus heterostrophus* (AAX_09989) nested within Bacterial Clade I by one or more strongly supported nodes (Figs. 1A and S1A). Both phylogeny estimates also place the Nested Fungal Clade (including *Byssochlamys*, *Penicillium*, and *Aspergillus* PKSs) within Bacterial Clade II (Fig. 1A, PPS = 100; Fig. S1A, BSS < 50). Additionally, two PKSs from the protozoan *Cryptosporidium parvum* (Eukaryota: Alveolata: Apicomplexa) formed a clade that was deeply phylogenetically nested within Bacterial Group II, and distantly related to the Nested Fungal Clade, the next most closely related eukaryotic PKSs (Figs. 1A and S1A).

3.3. Results of phylogenetic analyses based on the KS domain only

In a vast majority of cases, the phylogeny estimates based solely on the KS domain were very similar to the estimates based on the KS+AT domains (BMCMC-KS+AT and MP-KS+AT hereafter), but with comparatively less topology resolution

and lower support for relationships, consistent with the smaller number of characters in the KS-only alignment. The broad congruence between KS and KS+AT domain phylogeny estimates provides confirmation that, in general, the evolutionary history of the two domains appears to be congruent or shared, suggesting that estimates based on the KS+AT domains provide

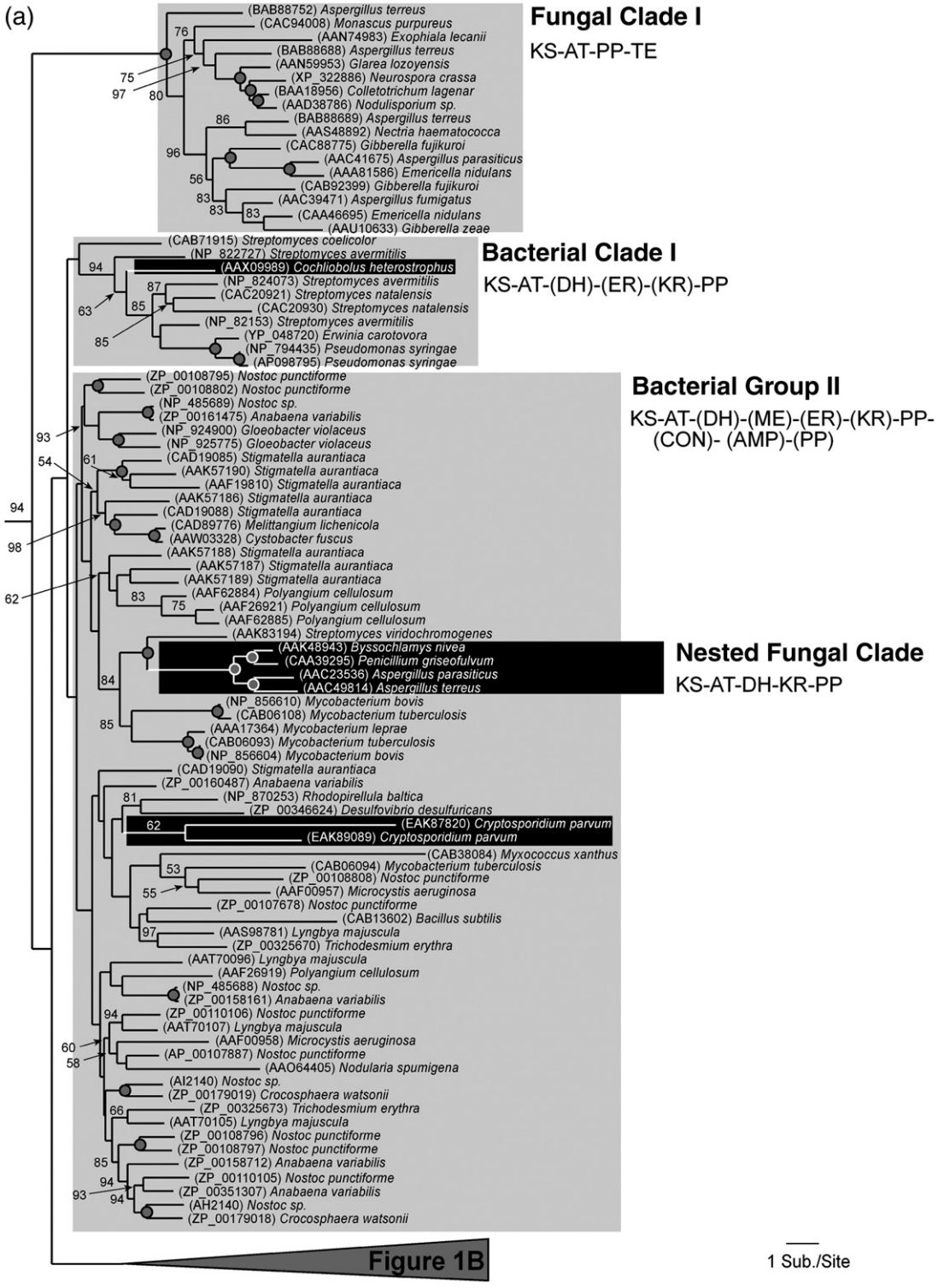

Fig. 1. Majority-rule consensus Bayesian phylogenetic estimate of relationships among type I polyketide and fatty acid synthases, based on an alignment of 552 amino acid positions (outgroups not shown). Posterior probability (PPS) values of nodal support > 49% are shown on the tree; nodes receiving 100% PPS are indicated with a gray-filled circle. Major clades or groups, indicated with gray-filled rectangles, are identified and characteristic protein structural domains for each clade are given. For domain annotations, domains in parentheses occur in only some members of particular groups or clades. See text (Section 2.1, or abbreviation section in front matter) for abbreviations used for protein domains. Black rectangles and white branches are used to indicate lineages or clades that are phylogenetically placed in an unexpected or unusual manner (potentially indicating horizontal gene transfer or differential gene loss).

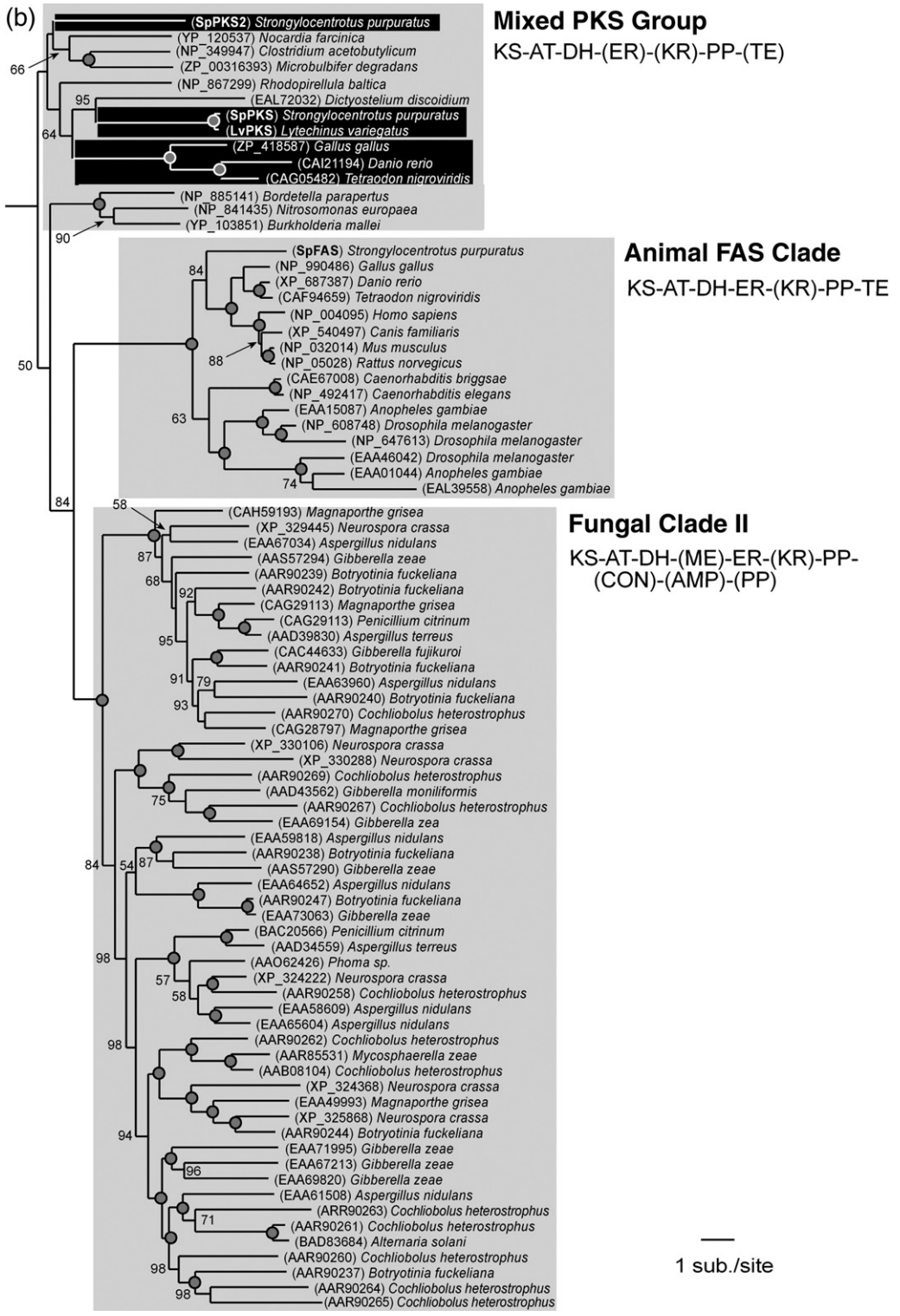

Fig. 1 (*continued*).

reliable, and likely superior, inferences of evolutionary history. Based on this result, we treat the inferred KS+AT domain phylogeny (particularly the BMCMC) as our preferred estimate and only focus on relevant relationships that were different between the KS and KS+AT phylogenies.

The MP estimate based on the KS-only alignment (MP-KS) was particularly poorly resolved and weakly supported, and we do not show these results for this reason. The entire BMCMC tree estimate based on the KS domain (BMCMC-KS) is provided as supplementary data (Fig. S2), and the summary of

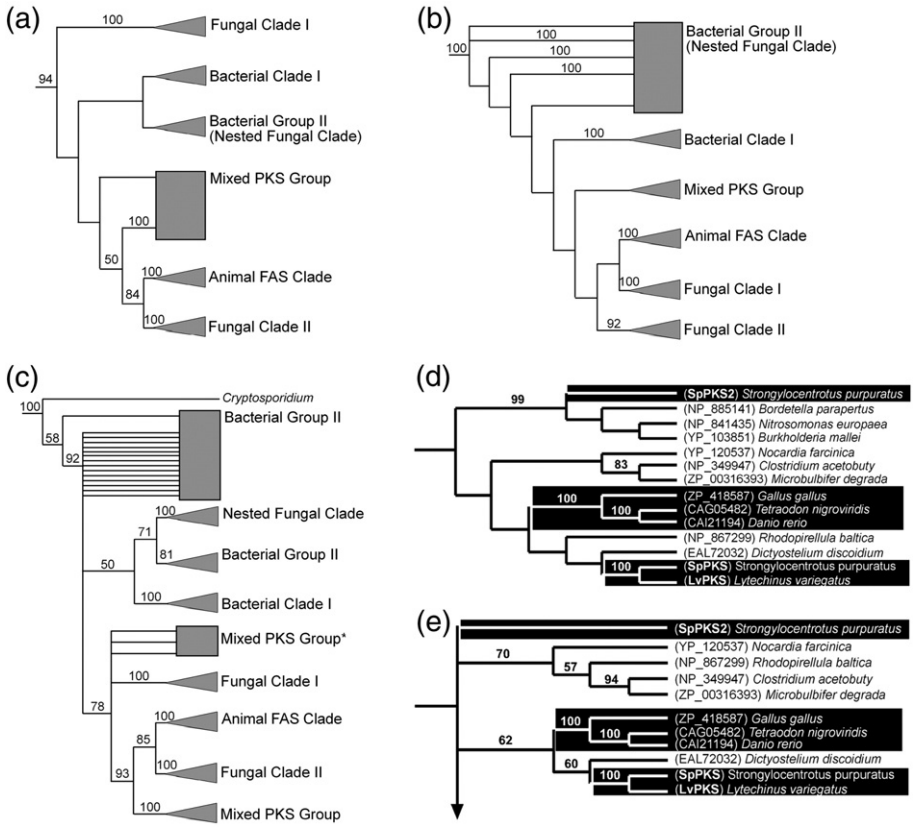

Fig. 2. Comparison between the branching order of major nominal clades or groups estimated by Bayesian (BMCMC; a) and maximum parsimony (MP; b) phylogenetic analyses of the KS and AT domain combined, and the BMCMC analysis of the KS domain only (c; outgroups not shown). The nomenclature for clades and groups follow Fig. 1. Nodal support for relationships among major clades is indicated (representing posterior probabilities for BMCMC, and bootstrap values for MP). The MP estimate of relationships among members of the Mixed PKS Group using the KS+AT domain alignment (d), and the BMCMC estimate based on the KS domain only (e), to compare with the BMCMC estimate in Fig. 1B. (e) corresponds with the grouping labeled "Mixed PKS Group*" in (c). Black-shaded rectangles (with white branches and text) indicate polyketide synthases identified from animal genomes.

higher-level relationships based on this analysis is provided for comparison (Fig. 2c).

The BMCMC-KS estimate of major relationships (Fig. 2c) is similar to the MP-KS+AT estimate (Fig. 2b) in inferring an early divergence of Bacterial Group II proteins, and a closer relationship of the two large fungal clades. Like the BMCMC-KS+AT estimate, however, the BMCMC-KS estimate implied a sister group relationship between the Animal FAS Clade and the Fungal Clade II (Fig. 2). As in all other estimates, the Nested Fungal Clade was placed within Bacterial Group II, but the BMCMC-KS estimate placed this grouping as the sister clade to Bacterial Group I PKSs (Fig. 2c). Like the MP-KS+AT, the Bacterial Group II PKSs did not form a clade, although support values were low (Fig. 2c, see also Fig. S2).

Consistent with all other estimates (including the MP-KS), the BMCMC-KS estimate inferred that PKSs and FASs from animal genomes formed distantly related clades, separated by non-animal proteins (Fig. 2c). All non-FAS animal proteins fell within the group labeled "Mixed PKS Group*" in Fig. 2c, which is illustrated in detail in Fig. 2e. As in the BMCMC-KS+AT tree, the BMCMC-KS phylogeny implies that the closest relative of the sea urchin proteins SpPKS and LvPKS is a PKS from the slime mold *Dictyostelium*, and that the sister group of this clade is a cluster of PKSs from the chicken (*Gallus*) and fish (*Danio* and *Tetraodon*) genomes (Fig. 2e). The phylogenetic placement of a second sea urchin PKS (SpPKS2, from *Strongylocentrotus*) was not well-supported, but was generally placed close to this animal PKS grouping in the BMCMC-KS estimate (Fig. 2e). The exact topology surrounding these animal PKSs was not well-supported, but included several lineages of bacterial and fungal PKSs (Fig. 2c, e, see also Fig. S2).

Like the KS+AT-based phylogeny estimates, the BMCMC-KS tree placed several fungal and protozoan PKSs nested within larger clades of bacterial PKSs (see black boxes in Fig. S2). As inferred in the MP-KS+AT and BMCMC-KS+AT estimates, the BMCMC-KS analysis placed the Nested Fungal Clade among bacterial PKSs, although it was estimated as the sister lineage to a cluster of Bacterial Group II proteins (Fig. 2c). Also, the two PKSs from the protozoan *C. parvum* formed a clade deeply phylogenetically nested within Bacterial Group II, as in the MP-KS+AT and BMCMC-KS+AT trees (Fig. S2). The only major difference regarding the topology of apparently misplaced PKSs between KS and KS+AT datasets was the placement of the PKS from the fungus *C. heterostrophus*. The BMCMC-KS analysis inferred this protein to be deeply divergent based on the KS domain, whereas it was deeply nested within Bacterial Clade I in both KS+AT estimates (Figs. 1A, S1A, and 2c; see also Fig. S2).

4. Discussion

4.1. Enigmatic origins of a novel group of animal (non-fas) pks genes

The primary goal of this study was to investigate the phylogenetic placement (and potential orthologs) of the sea urchin *pks* gene *SpPks*. Early indications, during the first characterization of this gene (Calestani et al., 2003), suggested a complex evolutionary origin and a lack of animal orthologs. Results of our phylogenetic analyses suggest this gene (and what appears to be an ortholog [*LvPks*] from another sea urchin species, *L. variegatus*) is most closely related to a *pks* from the slime mold *Dictyostelium discoidium* (Figs. 1B and 2d, e). Despite intense searching of other animal genomic databases, we found no protein sequences in available animal (or fungal) genomes that were inferred as the closest relative of these sea urchin genes. We did, however, find a second clade of non-FAS PKSs (from chicken and fish genomes) that was closely related to SpPKS, but not more closely related than the *Dictyostelium* PKS. Furthermore, we found an additional predicted protein from the sea urchin *S. purpuratus* (SpPKS2; representing a distinct locus) that was inferred to be a member of this Mixed PKS Group, phylogenetically clustered with bacterial and/or fungal PKSs, although no BMCMC analysis strongly supported the precise placement of this protein (Figs. 1B and 2d, e).

In addition to the enigmatic evolutionary origins of *SpPks* and *SpPks2*, the results of phylogenetic analyses provide a similarly peculiar result in that chicken and fish genomes also contain a *pks* closely related to *SpPks*. None of these genes in chicken or fish have been characterized at the molecular level, and all represent predicted proteins. Of these vertebrate *pks* genes, expression (transcription) has only been confirmed for the chicken *pks* gene, based on cDNA clone libraries obtained from chicken macrophages and lymphocytes. These three lineages or clusters of animal *pks* genes placed within the Mixed PKS Group are clearly distantly related to *fas* genes of animals, and appear to have an evolutionary origin unique from animal *fas*.

To evaluate the possible evolutionary mechanisms (i.e., gene duplication/loss or HGT) that may have lead to the establishment of (non-*fas*) animal *pks* genes, it is critical to consider the underlying organismal phylogeny. Unfortunately, our understanding of the tree of life is still somewhat vague and in a process of refinement. Recent phylogenomic evidence (Ciccarelli et al., 2006) suggests that slime molds (i.e., *Dictyostelium*) are the sister lineage to animals and fungi (collectively). To avoid evoking HGT to explain the close relationship between sea urchin and *Discyostelium* PKSs, massive gene loss would need to have occurred in fungi, ecdysozoan animals (insects and nematodes), and in vertebrates. If our topology estimates are inaccurate, and instead the *Dictyostelium* PKS is the sister lineage to sea urchin, chicken, and fish PKSs, massive differential gene loss in even more lineages would be required to explain these phylogenomic patterns. Although HGT may seem a more parsimonious hypothesis than *pks* gene loss in multiple animal and fungal lineages, evidence that gene loss on similar scales may be somewhat common is accumulating in the literature. Evidence that fungi are particularly prone to *pks* gene duplication and loss (Kroken et al., 2003), and that substantial gene loss has differentially occurred across vertebrate (e.g., Dachin et al., 2005; Blomme et al., 2006) and invertebrate lineages (e.g., Hughes and Friedman, 2005), collectively imply that gene loss may be a reasonable explanation for the enigmatic phylogenomics of animal (non-*fas*) *pks* genes. Ongoing genome and EST sequencing projects in early diverging animal lineages (e.g., hemichordates, other echinoderms, mollusks, annelids) should eventually provide further insight into what other animal genomes may also possess orthologs to *SpPks* or the other animal non-*fas pks* genes (within the Mixed PKS Group).

Another interesting layer of evolutionary complexity regarding members of this Mixed PKS Clade has recently come to light. The *Dictyostelium* protein (steely) inferred to be the closest relative to SpPKS (and LvPKS) actually appears to be a hybrid (or fused) PKS, nominally referred to as a "hybrid type I fatty acid-type III polyketide synthase" (Austin et al., 2006). Our phylogenetic results suggest, instead, that at least the KS and AT domains of the *Dictyostelium pks* evolved from a non-*fas pks* type I ancestor, while the C-terminal domains, including a chalcone synthesase domain (CHS; normally present only in type III PKSs), may have been independently evolutionary acquired from a type III PKS. This *Dictyostelium* PKS produces a signaling molecule that modulates cell type differentiation. Given the drastic differences in the C-terminal domains, it is unlikely that the *Dictyostelium* steely PKS is functionally similar to the closely related animal PKSs.

4.2. Functionality of animal (non-FAS) PKSs

SpPks is expressed in the pigmented cells embedded in the epithelium of the larva and it is required to produce the pigment echinochrome (Calestani et al., 2003). The role of pigment cells is unclear, but they have been suggested as being involved in microbial defense (which is consistent with the functionality of polyketide compounds). In support of the hypothesis that SpPKS may play a role in immuno-defense, Service and Wardlaw (1984) purified echinochrome A from coelomic fluid (coelomocytes) of adult sea urchins, and showed that this compound possessed antibiotic properties against various marine and non-marine bacteria. Echinochrome A is also synthesized during the larval stage in pigment cells. The morphology and behavior of pigment cells are, to some extent, similar to that of macrophages. Pigment cells have a stellate shape with 2–3 pseudopodia, which can be rapidly extended and contracted, and they have the ability to migrate within the larval epithelium and the basal lamina (Gibson and Burke, 1987). Nevertheless, a direct involvement of pigment cells in the immuno-defense of the larva has yet to be shown. Evidence from chicken EST data does not provide a direct functional characterization of the chicken (non-FAS) PKS, although the results do provide some indications of its potential functionality. The only chicken EST libraries that have shown expression of this gene have been from macrophages and lymphocytes, also suggesting a potential antimicrobial role for the chicken (non-FAS) PKS.

4.3. Phylogenetic evidence for potential HGT and/or massive loss of pks genes in multiple eukaryotic lineages

Many PKSs, particularly in bacteria and fungi, synthesize polyketide compounds that clearly convey an important functionality (e.g., roles in microbial defense) and are likely to be subjected to intense evolutionary selection. Particularly high numbers of *pks* genes are found in some bacterial and fungal genomes (over twenty per genome in some eubacteria and fungi; Kroken et al., 2003; Jenke-Kodama et al., 2005), yet there are many notable contrasts between the genomes of related species in the number and diversity of *pks* genes present (e.g., Hopwood, 1997; Kroken et al., 2003; Jenke-Kodama et al., 2005). Collectively, these patterns suggest differential selection for (or against) the production of polyketide compounds drives an intense and complex evolutionary system of differential *pks* gene duplication and loss across lineages. This trend is clearly observed in ascomycete fungi, where no clear orthologous clusters of PKSs can be observed through the phylogenetic 'noise' of apparent differential gene duplication and loss (Kroken et al., 2003). In this context, it is not surprising that, in addition to this pattern of gene birth–death evolution, we may also observe successful HGT events. The horizontal transfer of genes is typically viewed as evolutionary unfavorable and selected against. The argument has been made that transferred genes should be lost to random mutational process, and will be stably maintained only under positive selection, whereby they provide a functional adaptive advantage (Kurland et al., 2003). *Pks* genes are typically expressed only under specific cellular circumstances or selective pressures and do not modify primary metabolism (Ginolhac et al., 2005), and experimental induction of stable and functional *pks* HGT is readily accomplished in bacteria (e.g., Shah et al., 2000; Volchegursky et al., 2000). These characteristics, together with the potential selective advantages conveyed by *pks* genes makes them ideal candidates for differential gene duplication and loss, as well as stable HGT.

Previous studies examining the *pks* gene family have repeatedly found evidence for unique instances of HGT. Many studies have pointed out compelling evidence for the HGT of type I *pks* genes between eubacterial lineages (Egan et al., 2001; Anzai et al., 2003; Jenke-Kodama et al., 2005). Actinomycete eubacteria, in particular, appear to have repeatedly undergone HGT of type I *pks* genes, and it is thought that this widespread HGT among this group of bacteria is favored due to the linearity and instability of actinomycete chromosomes, associated with the large quantity of mobile genetic elements they contain (Ginolhac et al., 2005). Kroken et al. (2003) were the first to thoroughly examine the diversity of type I PKSs in fungi (focusing on ascomycetes) using a phylogenetic approach. Their results indicated two lineages of fungal PKS proteins that were phylogenetically placed apart from other fungal PKSs, and phylogenetically nested within clades of bacterial PKS proteins. They concluded that these phylogenetic results may indicate HGT of bacterial *pks* genes into fungal genomes. In this study, we have a substantially expanded sampling of bacterial PKSs, and we find similar relevant phylogenetic results, except that in our topologies these fungal lineages are more deeply nested within bacterial PKS clades.

The first of these cases implying potential HGT is the strongly supported nesting of the PKS from the fungus *C. heterostrophus* (AAX09989; also referred to as PKS24 or NPS7) within Bacterial Clade I, based on all (KS+AT and KS-only) phylogenetic estimates (Figs. 1A, S1A, and S2A). This *pks* from *C. heterostrophus* is unique in that it encodes a hybrid PKS/nonribosomal peptide predicted to synthesize a partly reduced polyketide decorated with a single amino acid (Kroken et al., 2003). The second instance of phylogenetically misplaced fungal PKSs (as also observed by Kroken et al., 2003) is nominally referred to as the Nested Fungal Clade, placed among Bacterial Group II proteins in all analyses (Figs. 1A, S1A, and S2A). Due to our increased sampling of bacterial PKSs, our results provide an interesting strengthened line of support for HGT from bacterial to fungal lineages in both cases. Our phylogenies estimate both of these misplaced fungal PKS lineages to be nested within clusters of Actinomycete (e.g., *Streptomyces* spp.) eubacterial PKSs, and are consistent with previous evidence suggesting elevated rates of type I PKS gene HGT in this group of bacteria (Ginolhac et al., 2005).

We found particularly interesting results regarding the phylogenetic placement of PKS proteins from the eukaryotic protozoan *C. parvum* (an intracellular parasite of vertebrates); this was the only strongly supported drastic change between KS+AT and KS-only phylogeny estimates. Phylogenies based on the KS+AT domains place these *Cryptosporidium* PKSs deeply nested among eubacterial lineages within Bacterial Group II (Figs. 1A and S1A). Contrastingly, the BMCMC-KS phylogeny estimate placed these *Cryptosporidium* PKSs at the base of the type I PKS phylogeny. We interpret these results as implying that the KS and AT domain of these PKSs most likely have substantially different evolutionary histories and origins, with the AT domain potentially stemming from a bacterial origin (i.e., near the placement in the KS+AT trees), whereas the KS domain is so distantly related to other type I PKS KS domains, it may not have even originated from a type I PKS. Thus, domain shuffling (across PKS types) and HGT may have evolutionarily given rise to these *Cryptosporidium pks* genes.

4.4. Conclusions

Our phylogenomic data suggest multiple evolutionary origins of animal *pks* and *fas* genes, with a single origin of animal *fas*, and a complex and hereto unknown evolutionary origin of several non-*fas pks* genes in animals. The enigmatic origins of non-*fas pks* genes in sea urchins, birds, and fish suggest a complex evolutionary scenario, most likely explained by massive *pks* gene loss in multiple animal (and other eukaryote) lineages leading to the occurrence of these genes in only some animal genomes. Additional comparative genomic data from diverse animal genomes will be necessary to more conclusively rule out HGT and evaluate the hypothesis of multi-lineage gene loss in the case of animal *pks* genes. Available evidence for the role of the polyketide products of these PKSs remains vague, although preliminary indications that these

PKSs may play roles in innate immunity (especially in sea urchins which are not known to have an adaptive immune system) present an intriguing potentially novel functional role for PKSs (and polyketides) in animals that requires more attention.

The finding that animal *fas* and the Fungal Clade II (based on BMCMC, or even both major fungal clades based on MP) may comprise sister evolutionary lineages highlights a dramatic example of strong differential gene family evolution stemming from a single *pks* gene in the ancestor of animals and fungi (this Fungal Clade II is representative of the fungal reducing PKS group of Kroken et al., 2003). Whereas animal genomes have not undergone any substantial duplication or neofunctionalization of this ancestral locus, fungal genomes appear to have undergone an intense process of gene duplication and loss, coupled with differential accessory domain arrangements to produce a broad diversity of PKS proteins (and polyketide compounds). We found evidence for two large groups of fungal type I PKSs (Fungal Clades I and II), and BMCMC results suggest that these two groups may have different evolutionary origins. Our results also support two other independent origins of fungal PKSs, and an anomalous origin of protozoan PKSs, that (in accordance with previous hypotheses) may have resulted from HGT.

Establishing conclusive evidence for HGT is difficult and controversial (e.g., Kurland et al., 2003), particularly given accumulating evidence for previously unappreciated high levels of differential gene loss across eukaryotic lineages. Within the constraints posed by the current genomic data available, it appears that HGT may be prevalent and significant in the evolutionary history of type I *pks* genes across broad taxonomic groups, including eubacteria, protozoans, fungi, although probably less so in animals. Despite the current limited comparative genomic data, our results together with previous studies demonstrate that differential gene family expansion and contraction, due to differential gene duplication and loss across lineages, is an obvious and strong component of the evolution of the type I PKS gene family. Collectively, the complex phylogenomics associated with the *pks* gene family are likely related to the particularly significant functionality (i.e., antimicrobial, predator defense, etc.) conveyed by the polyketide products synthesized by PKSs. Continuing expansion of genomic resources for diverse taxa will be particularly important for clarifying the relative roles of HGT and gene duplication/loss that have led to the diversity of *pks* genes (and polyketide compounds) across domains of life. These data, coupled with additional work to characterize the functional roles of the polyketide products produced by this gene family, would provide a biologically fascinating example of gene family functional evolution, while also providing an extremely useful biopharmaceutical resource.

Acknowledgments

We thank Jill Castoe and Robert Ruggiero for constructive comments on various drafts of the manuscript. Scott Kroken provided significant assistance with the identification and understanding of PKS domain nomenclature. We thank John Gerhart for providing access to the *Saccoglossus kowalevskii* EST collection. We thank Christopher Parkinson, Patrick Keeling, and Kevin Peterson for providing valuable advice on the project, and Christopher Parkinson for providing computational resources purchased with funds from an NSF grant (DEB-0416000) and a UCF startup package. CC was supported by a UCF startup package.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.gene.2006.11.005.

References

Anzai, Y., Saito, N., Tanaka, M., Kinoshita, K., Koyama, Y., Kato, K., 2003. Organization of the biosynthetic gene cluster for the polyketide macrolide mycinaminicin in *Micromonospora griseorubida*. FEMS Microbiol. Lett. 218, 135–341.

Austin, M.B., et al., 2006. Biosynthesis of Dictyostelium discoideum differentiation-inducing factor by a hybrid type I fatty acid-type III polyketide synthase. Nat. Chem. Biol. 2, 494–502.

Blomme, T., Vanderpoele, K., De Bodt, S., Simillion, C., Maere, S., Van de Peer, Y., 2006. The gain and loss of genes during 600 million years of vertebrate evolution. Genome Biol. 7, R43.

Calestani, C., Rast, J.P., Davidson, E.H., 2003. Isolation of pigment cell specific genes in the sea urchin embryo by differential microarray screening. Development 130, 4587–4596.

Castoe, T.A., Doan, T.M., Parkinson, C.L., 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. Syst. Biol. 53, 448–469.

Cicarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B., Bork, P., 2006. Toward automatic reconstruction of a highly resolved tree of life. Science 311, 1283–1287.

Dachin, E.G.J., Gouret, P., Pontarotti, P., 2005. Eleven ancestral gene families lost in mammals and vertebrates while otherwise universally conserved in animals. BMC Evol. Biol. 6, 5.

Egan, S., Wiener, P., Kallifidas, D., Wellington, E.M., 2001. Phylogeny of *Streptomyces* species and evidence for horizontal gene transfer of entire and partial antibiotic gene clusters. Antonie Van Leeuwenhoek 79, 127–133.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Garson, M.J., 1989. Biosynthetic studies on marine natural products. Nat. Prod. Rep. 6, 143–170.

Gibson, A.W., Burke, R.D., 1987. Migratory and invasive behavior of pigment cells in normal and animalized sea urchin embryos. Exp. Cell Res. 173, 546–557.

Ginolhac, A., et al., 2005. Type I polyketide synthases may have evolved through horizontal transfer. J. Mol. Evol. 60, 716–725.

Griffiths, M., 1965. A study of the synthesis of naphthaquinone pigments by the larvae of two species of sea urchins and their reciprocal hybrids. Dev. Biol. 11, 433–447.

Gross, F., et al., 2006. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. Arch. Microbiol. 185, 28–38.

Hopwood, D.A., 1997. Genetic contributions to understanding polyketide synthases. Chem. Rev. 97, 2465–2497.

Hopwood, D.A., Sherman, D.H., 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. 24, 37–66.

Huelsenbeck, J.P., 1995. The performance of phylogenetic methods in simulation. Syst. Biol. 44, 17–48.

Huelsenbeck, J.P., Crandall, K.A., 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. Annu. Rev. Ecol. Syst. 28, 437–466.

Hughes, A.L., Friedman, R., 2005. Loss of ancestral genes in the genomic evolution of *Ciona intestinalis*. Evol. Dev. 7, 196–200.

Jenke-Kodama, H., Sandmann, A., Müller, R., Dittman, E., 2005. Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. 22, 2027–2039.

Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., Turgeon, B.G., 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc. Natl. Acad. Sci. U. S. A. 100, 15670–15675.

Kuhn, R., Wallenfells, K., 1940. Echinochrome als prosthetische gruppen hochmolekularern symplexe in den eiern von *Arbacia pustulosa*. Ber. Dtsch. Chem. Ges. 73, 458–464.

Kurland, C.G., Canbank, B., Berg, O.G., 2003. Horizontal gene transfer: a critical view. Proc. Natl. Acad. Sci. U. S. A. 100, 9658–9662.

Mar, J.C., Harlow, T.J., Ragan, M.A., 2005. Bayesian and maximum likelihood phylogenetic analyses of protein sequence data under relative branch-length differences and model violation. BMC Evol. Biol. 5, 8.

McLendon, J.F., 1912. Echinochrome, a red substance in sea urchins. J. Biol. Chem. 11, 435–441.

Moore, B.S., Hopke, J.N., 2001. Discovery of a new bacterial polyketide biosynthetic pathway. ChemBioChem. 2, 35–38.

Moss, S.J., Martin, C.J., Wilkinson, B., 2004. Loss of co-linearity by modular polyketide synthases: a mechanism for the evolution of chemical diversity. Nat. Prod. Rep. 21, 575–593.

Rambaut, A., Drummond, A.J., 2003. Tracer, Version 1.0.1. Available from http://evolve.zoo.ox.ac.uk/.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Salaque, A., Barbier, M., Lederer, E., 1967. Sur la biosynthèse de l'échinochrome A par l'oursin Arbacia pustulosa. Bull. Soc. Chim. Biol. 49, 841–848.

Schroder, J., et al., 1998. Plant polyketide synthases: a chalcone synthase-type enzyme which performs a condensation reaction with methylmalonyl-CoA in the biosynthesis of C-methylated chalcones. Biochem. 37, 8417–8425.

Service, M., Wardlaw, A.C., 1984. Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L.). Comp. Biochem. 79B, 161–165.

Shah, S., Xue, Q., Tang, L., Carney, J.R., Batlach, M., McDaniel, R., 2000. Cloning, characterization and heterologous expression of a polyketide synthase and P-450 oxidase involved in the biosynthesis of the antibiotic oleandomycin. J. Antibiot. 53, 502–508.

Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18, 380–416.

Volchegursky, Y., Hu, Z., Katz, L., McDaniel, R., 2000. Biosynthesis of the anti-parasitic agent megalomicin: transformation of erythromycin to megalomicin in *Saccharopolyspora erythraea*. Mol. Microbiol. 37, 752–762.

Whelan, S., Goldman, N., 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol. Biol. Evol. 18, 691–699.

Winkel-Shirley, B., 2002. Biosynthesis of flavonoids and effects of stress. Curr. Opin. Plant Biol. 5, 218–223.