Wheels within Wheels: Clues to the Evolution of the *Gnas* and *Gnal* Loci

Samir Wadhawan,¹ Benjamin Dickins, and Anton Nekrutenko¹

Center for Comparative Genomics and Bioinformatics, Huck Institutes for the Life Sciences, Penn State University

The *Gnas* and *Gnal* loci, which encode the α subunits of stimulatory G-proteins, are among the most complex eukaryotic genes. They combine elaborate patterns of imprinting, alternative splicing, and antisense transcription with tissue- and developmental stage-specific expression. Different regions of these genes evolve at drastically different rates such that some show complete conservation, whereas others are virtually unalignable. Yet, the most unusual feature of the *Gnas/Gnal* complex is the presence of the longest known overlap between coding regions resulting in the production of two unrelated proteins: XLs and its putative regulator ALEX. Here we elucidate the evolutionary history of both loci and uncover new complexities. First, alternatively spliced regions of both loci evolve under varying selective regimes echoing their distinct biological roles. Second, an enigmatic alternative transcript of the *Gnas* locus, known as Nesp, is likely bicistronic. Third, rodent XLs and ALEX follow an evolutionary trajectory distinct from that of other mammals and show extensive sequence variation in the internal repeat region, a fact that might be explained by variation in the robustness of imprinting. Fourth, we show that the overlap between the XLs and ALEX frames is restricted to eutherian mammals. Finally, we reconcile our findings with extensive physiological data derived from animal models.

Introduction

Guanine nucleotide-binding proteins or G-proteins are molecular transistors that are individuated on the basis of variation in one of their three component subunits: the α (alpha) subunit. Here we focus on the evolution of two highly complex paralogous genes encoding vertebrate α subunits, namely *Gnas* and *Gnal*. The mouse *Gnas* locus on distal chromosome 2 (like its ortholog, *GNAS*, on human 20q13) is characterized by a bewilderingly complex pattern of splicing in which multiple transcripts are produced from alternative first exons spliced principally to shared downstream exons. The *Gnas* locus also shows tissue-specific and imprinted expression such that different transcripts are expressed in different tissues and from maternal versus paternal copies. In total, at least four functionally distinct proteins are produced (fig. 1) (Plagge and Kelsey 2006), and of the locus’s five promoters, only one drives biallelic transcription. This sole promoter directs expression of the canonical locus transcript, *Gnas*, in many tissues although silencing of its paternal copy occurs in a subset of tissues (Yu et al. 1998; Hayward et al. 2001; Germain-Lee et al. 2002; Mantovani et al. 2002; Liu et al. 2003). The remaining four promoters overlap with three differentially methylated regions (DMRs) and drive expression of maternal- or paternal-specific transcripts (Li et al. 2000). Of these four promoters, two produce exclusively noncoding paternally derived transcripts and one of these is in the antisense orientation (Ischia et al. 1997; Hayward, Kamiya, et al. 1998; Hayward, Moran, et al. 1998; Hayward and Bonthron 2000; Wroe et al. 2000).

Among the proteins encoded by the *Gnas* locus are several variants of the classical stimulatory G-protein alpha subunit Gzs as well as two others that share no sequence homology with Gzs: Nesp55 and ALEX (fig. 1). Nesp55 is encoded by its own large upstream exon spliced to Gzs exons 2–13. Results of in vitro experiments suggest that Nesp55 may be bicistronic producing Nesp55 itself and a truncated version of Gzs (Ishikawa et al. 1990). Yet, it is the encoding of ALEX that sets *Gnas* apart from any other gene: ALEX’s entire reading frame is embedded within the *Gnasxl* first exon (herein the XL exon of *Gnas*; fig. 1). At over 1,000 bp, it is, to the best of our knowledge, the longest overlap between two protein-coding regions ever found in eukaryotic genomes (Chung et al. 2007). In the *Gnasxl* transcript, the XL exon is contiguous with downstream exons 2–13 of *Gnas* and the principal reading frame beginning in the XL exon stays in frame with these and gives rise to a Gzs variant called XLs (Kehlenbach et al. 1994; Ischia et al. 1997; Klemke et al. 2001). Although the XLs frame continues to downstream exons, the ALEX frame terminates exactly at the end of XL exon (Klemke et al. 2001). The extensive overlap between the XLs and ALEX reading frames is therefore confined to the unusually large XL exon (1,300 bp in human and 1,447 bp in mouse). Despite the fact that both reading frames are preserved in examined primates and rodents, their XL exons are divergent to the extent of being virtually unalignable (Klemke et al. 2001).

The two reading frames within the XL exon are in the same orientation but shifted one nucleotide relative to each other so that codon positions 1, 2, and 3 of the first frame overlap with positions 3, 1, and 2 of the second frame. Both the XLs and ALEX proteins are produced from the *Gnasxl* transcript from alternative translation initiation starts (Klemke et al. 2001). Furthermore, the two proteins interact with one another (Klemke et al. 2001; Freson et al. 2003). XLs mediates signal transduction, in the same manner as Gzs, by linking activation of G-protein-coupled receptors with activation of adenyl cyclase (Klemke et al. 2000; Bastpe et al. 2002). The function of the ALEX polypeptide is less clear, but it appears to regulate the signaling properties of XLs. Binding of ALEX to XLs is hypothesized to prevent the latter from interacting with receptors (Freson et al. 2001, 2003).

*Gnal* is paralogous to *Gnas* and encodes the olfactory specific stimulatory G-protein alpha subunit Golf (Vuuristo et al. 2000). It is the second member of the Gs subfamily of G-proteins in vertebrates. The two loci are similarly structured, and *Gnal* possesses a large upstream exon called 1a (fig. 1) that is paralogous to the XL exon of *Gnas* (Corradi et al. 2005) (We use the naming convention depicted in fig. 1 throughout this manuscript). The *Gnal* transcript beginning with the XL exon encodes the

¹ These authors contributed equally.

Key words: overlapping reading frames, imprinting, protein-coding regions.

E-mail: anton@bx.psu.edu.

doi:10.1093/molbev/msn229
Advance Access publication October 8, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
extra large variant of the canonical Gaolf known as XLGaolf but lacks the alternative reading frame found in the XL exon of Gnas. The XL exon of Gnal does, however, overlap with a DMR and may therefore be subject to imprinting (Vuoristo et al. 2000). There is no evidence for the existence of a paralog of the Nesp55-coding exon or for antisense transcription in the Gnal locus.

Gnas and Gnal are not the only genes that code for G-protein alpha subunits. The Ga superfamily, which incorporates the Gs subfamily, has evolved through a complex series of whole gene and exon duplications (Wilkie et al. 1992). In vertebrates, there are 15 Ga genes distributed among four subfamilies, Gs, Gq, Gi, and G12, that can be distinguished based on between-member sequence homology, receptor interactions, and coupling to downstream effectors (reviewed by Neves et al. 2002). Gzs and Gaolf, the Gs subfamily, stimulate specific adenylyl cyclases, which, in turn, catalyze the production of the second messenger cAMP from ATP. Gzs is also known to activate Src-kinase and cardiac L-type calcium channels as well as to inhibit cardiac sodium channels (reviewed in Morris and Malbon [1999]). Activated subunits from the Gq family (Gqz, Gz11, Gz14, and Gz16) stimulate phospholipase C to produce the intracellular messengers inositol triphosphate (IP3) and diacylglycerol, which regulate calcium homeostasis (Sternweis and Smrcka 1992). Signaling cascades initiated by members of the Gi and G12 families have not been well defined, but the Gi family (Gzi1-3, Gzi1-2, Gzgust, Gzo, and Gzz) is mostly studied for its role in the transducin pathway, which mediates light detection in the eye (Pfeuffer and Helmreich 1988). Subunits of the G12 family (Gz12 and Gz13) couple with the lysosphatidic acid and thromboxane A2 receptors (Offermanns et al. 1994; Gohla et al. 1998), and their target is phospholipase D (Singh et al. 2005). Because genes from all four subfamilies are found in Drosophila, they are thought to have evolved prior to the divergence of vertebrates and invertebrates (Wilkie et al. 1992).

The remarkable complexity of the Gnas and Gnal loci raises interesting questions we attempt to address here. First, as transcripts of the two genes possess distinct functions, is there evidence of differential selective constraints acting upon alternatively spliced upstream exons in different species? Second, the XL exon, which contains the overlap between the Gnasxl and Alex reading frames, evolves at an unprecedented rate with mouse and rat showing the highest divergence relative to other mammals. Is this trend common to all rodents and what are the biological implications of this? Finally, what sequence of events lead to the evolution of the two loci and emergence of dual coding within Gnas?
Materials and Methods
Amplification and Sequencing of XL Exon

Tissue samples for Apodemus agrarius, Apodemus sylvaticus, Mastomys huberti, Prazomys fumatus, Mus caroli, Mus hortulanus, Mus pahari, and Mus spretus were kindly provided by Dr. Robert J. Baker (Department of Biological Sciences and the Museum, Texas Tech University). The XL exon was amplified from all species using primers Mm-433F and Mm-2109R (supplementary table S1, Supplementary Material online). Mus species were sequenced using “Mseq” primers, while remaining rodents were sequenced using species-specific primers (supplementary table S1, Supplementary Material online). The remaining primers were designed using the Mus musculus XL exon sequence obtained from the University of California at Santa Cruz Genome Browser (http://genome.ucsc.edu). Polymerase chain reaction (PCR) was performed using 1.75 U Taq (Expand High Fidelity PCR System, Roche Diagnostics, Pleasanton, CA), 0.2 mM deoxynucleoside triphosphates, 300 nM of each primer, 50 ng/µl template DNA, PCR buffer with MgCl2 (Expand High Fidelity PCR System), and 7% dimethyl sulfoxide (DMSO). Hot start reaction was carried out using an ABI Thermocycler 9700 under the following conditions: 94°C for 5 min (initial denaturation) with consequent 30 cycles of denaturation at 94°C for 30 s, and annealing at 61°C for 30 s, elongation at 72°C for 2 min, followed by a final extension at 72°C for 5 min. The amplified products were purified using the QIAquick PCR purification kit (Qiagen Inc., Valencia, CA). Sequencing reactions were carried out using 1 µM of primers, 7% DMSO, 35–50 fmol of template DNA, and CEQ DTCS Quick Start Kit (Beckman Coulter, Fullerton, CA) in an ABI Thermocycler 9700 under the following conditions: 40 cycles of 96°C for 20 s, 50°C for 20 s, and 60°C for 4 min. Traces were obtained using Beckman Coulter CEQ8000 sequencer. Sequence traces were manually analyzed using the DNAStar software package (http://www.dnastar.com/) and deposited to GenBank under the following accession numbers: EF432823 (A. agrarius), EF432824 (A. sylvaticus), EF432825 (M. huberti), EF432826 (P. fumatus), EF432827 (M. caroli), EF432828 (M. hortulanus), EF432829 (M. pahari), and EF432830 (M. spretus).

Estimation of Nucleotide Substitution Rates

Sequences corresponding to the coding regions of Nesp55, Gnas xL, Gnas 1, Gnas 2–13, Gnal xL, Gnal 1, and Gnal 2–13 (fig. 1) were obtained for Homo sapiens (human), Macaca mulatta (macaque), Canis familiaris (dog), Bos taurus (cow), M. musculus (mouse), and Rattus norvegicus (rat) using the UCSC Genome Browser (http://genome.ucsc.edu) and ENSEMBL (http://www.ensembl.org). The protein sequences were aligned using MUSCLE (Edgar 2004). These alignments were used as a guide to realign corresponding nucleotide sequences. The dN–dS estimation, parametric bootstrap, transition/transversion ratio calculation, and tree reconstruction were carried out with HYPHY package (Pond et al. 2005).

Definition of Amino Acid Repeats Encoded by XL Exon

We used the HHrep Web server (Soding et al. 2006) to demarcate amino acid repeats within XLazs segments encoded by the XL exons of human, mouse, A. agrarius, and dog. The following parameters were used to produce the plots shown in figure 4: Sequence identity to query = 50, score threshold = 0.5, and window half width = 15.

Results

Gnas and Gnal: Conserved Arrangement of Divergent Parts

Gnas and Gnal produce a variety of transcripts with different expression patterns and distinct functions. Such functional plurality extends to several levels. First, each locus produces functionally and structurally distinct proteins. For example, Nesp55 and ALEX, both encoded by Gnas locus transcripts, are unrelated to each other and to the main product of the locus: Gzs. Second, different splice variants of the same transcript can produce structurally related but functionally distinct versions of the same protein as exemplified by the XLazs and XLGolf versions of the Gzs and Golf proteins, respectively. Such complexity might have been caused by distinct selective pressures acting upon different partitions of Gnas and Gnal. To test this possibility, we collected sequences of Gnas and Gnal orthologs from macaque, dog, cow, mouse, and rat genome assemblies. We separated the coding regions into partitions corresponding to the Nesp55 exon, the XL exon of Gnas, Gnas exon 1, Gnas exons 2 through 13, the XL exon of Gnal, Gnal exon 1, and Gnal exons 2 through 12 (fig. 1). For each partition, we constructed amino acid alignments of the encoded proteins (because the XL exon of Gnas is highly divergent only unambiguously aligning portions were used), our first objective being to obtain a bird’s eye assessment of divergence within these partitions. We assigned a pixel-wide bar to each amino acid (fig. 1): Residues identical to human were colored yellow and extended to maximum height, whereas discordant amino acids were colored brown and had progressively smaller height. The view obtained from these alignments illustrated differences in species divergences among the partitions (fig. 1). For example, Gnas exons 2–13 and Gnal exon 1 share >95% amino acid identity between human and mouse and >90% between human and cow. On the other hand, human NESPM55 is 75% identical in amino acid sequence to mouse Nesp55 and the XL exons of Gnas and Gnal exhibit <60% and <65% identity between the two species, respectively.

Does the variation in divergence levels among partitions of the two loci reflect differences in selection pressures? To answer this question, we compared dN/dS ratios (a standard evolutionary metric for protein-coding regions [Li 1997; Hurst 2002]) among the partitions using a modification of the Yang and Swanson (2002) approach implemented in the HYPHY package (Pond et al. 2005). For each comparison, we first fitted a codon substitution model (using the Muse and Gaut [1994] 94 model with additional parameters accounting for transition/transversion A/G and C/T transition biases) to each partition separately. Next, we refitted the model now fixing the transversion ratio calculation, and tree reconstruction were carried out with HYPHY package (Pond et al. 2005). The resulting ratio test was used to determine if the dN/dS ratio differs significantly between compared partitions. Results of this
Table 1
Results (*P* values) of Likelihood Ratio Test Contrasting Selective Pressures within and between *Gnas* and *Gnal* Partitions

<table>
<thead>
<tr>
<th></th>
<th><em>Gnas</em></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nesp</td>
<td>XL</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>XL</td>
<td>$5 \times 10^{-4}$</td>
<td>$3 \times 10^{-3}$</td>
<td>$8 \times 10^{-5}$</td>
<td>0.57</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>2–13</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th><em>Gnal</em></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>XL</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2–13</td>
<td>$5 \times 10^{-4}$</td>
<td>$1 \times 10^{-3}$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th><em>Gnas</em></th>
<th></th>
<th><em>Gnal</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>XL</td>
<td>$5 \times 10^{-8}$</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td>0.56</td>
</tr>
<tr>
<td>2–13</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

analysis are summarized in Table 1. In the *Gnas* locus, both the *Nesp55* and XL exons evolve significantly faster than exon 1 and exons 2 through 13 (note that because the XL exon contains two overlapping reading frames, the inferences of *dN* and *dS* may not be easily interpreted; this issue is addressed in detail below). Surprisingly, exon 1 of *Gnal* behaves very differently from its *Gnas* counterpart—its nonsynonymous rate is much higher and is similar to that of the XL exon of *Gnal*. A comparison of selective regimes between paralogous partitions in *Gnas* and *Gnal* (XL exon of *Gnas* vs. XL exon of *Gnal*, *Gnas* exon 1 vs. *Gnal* exon 1, and *Gnas* exons 2 through 13 vs. *Gnal* exons 2 through 12) shows that exons XL and 1 evolve at different rates in *Gnas* versus *Gnal*, whereas exons 2 through 13/12 follow a similar substitution trajectory in both loci.

The *Nesp55* exons are spliced to downstream *Gnas* exons to produce a variety of transcripts (fig. 1). The longest of these transcripts links the *Nesp55* exons with *Gnas* exons 2–13 and contains two reading frames. The first reading frame is contained entirely within *Nesp55* exons and codes for *Nesp55*. The start codon for the second reading frame (start A; fig. 2A) overlaps with the *Nesp55* stop and is in phase with Gzα protein–coding region within exons 2–13 (Weiss et al. 2000). If translated, the second reading frame would produce a truncated version of Gzα lacking N-terminal sequences (coded by exon 1 or by the XL exon in XLzα) that are required for interaction with β and γ G-protein subunits. The truncated Gzα reading frame also contains a second in-frame start codon (start B; fig. 2A) within *Gnal* exon 2. Although no published evidence points to in vivo translation of the truncated Gzα from the *Nesp55* transcript, a polypeptide initiated at start B within the exon 1A transcript (fig. 1) has been observed in vitro and in COS cell systems (Ishikawa et al. 1990). Thus, it is uncertain whether truncated Gzα is expressed from the *Nesp55* transcript or if it is biologically relevant. The availability of genomic sequences from multiple mammals allows us to examine this issue using the comparative genomics methodology. We

![Fig. 2.](image-url)
surveyed all available mammalian sequence data (including expressed sequence tags and partial genome assemblies) and generated an 11-way alignment of the Nesp55-coding exon that includes human, chimpanzee, macaque, cow, dog, cat, mouse, rat, pig, hedgehog, and elephant sequences (fig. 2B). The “TAATG” sequence serving as the stop of Nesp55 frame and as the start of the truncated Gzs frame is conserved in all species and invariably located 48 bp upstream of the 3’ boundary of the Nesp exon (fig. 2B). Even though such conservation is striking, it is not sufficient to establish the biological relevance of truncated Gzs. Because the Nesp55 and Gnas reading frames are shifted relative to each other within the Nesp55 exon, they should exhibit a periodicity characteristic of the protein-coding regions in their respective phases. Therefore, if the portion of the Gzs reading frame within Nesp exon possesses this property, it can be considered functional. A common approach for detecting such periodicity is to contrast the transition/transversion ratios (κ) of the first and second codon positions of a reading frame against κ at the third position (κ3). This approach is useful because, in most standard protein-coding regions, κ3 is significantly higher than the κ at the first and second codon positions (κ12) so that κ12 < κ3 (Li 1997). This is a consequence of the fact that most substitutions at the third codon position are synonymous, whereas all but eight substitutions are non synonymous in the first codon position and all are non synonymous in the second codon position. For the sake of the discussion, we refer to the section of the Nesp-coding exon occupied by the Nesp reading frame as the Nesp section and, that occupied by the Gnas frame, as the Gnas section (fig. 2). If the Gnas section encodes a functional protein, it must satisfy κ12 < κ3 only within Gzs frame, whereas this condition should not hold in the two remaining frames. Parametric bootstrap P values for testing of the condition that κ12 < κ3 (from 1,000 replicates) were 0.04, 0.39, and 0.56 for Gzs, +1, and +2 frames, respectively, suggesting that Gzs frame is likely functional. The P value of 0.04 is only marginally significant, presumably owing to the small size of the data set (48 sites in 11 species). But, combined with the observed high level of conservation, it makes the functionality of Nesp55-derived truncated Gzs highly likely.

Evolutionary Oscillation within XL Exon: Rodents Are Unique

The XL exon is the most unusual feature of the Gnas locus. It is large (~2,000 kb in humans; for comparison, the average length of human protein–coding exons is 177 bp), contains what we consider to be the longest dual-coding overlap between two protein-coding regions known in eukaryotic systems, and evolves at an unprecedented rate. Moreover, the products of the two overlapping reading frames interact in a highly specific manner (Freson et al. 2003). Our previous study in primates (Nekrutenko et al. 2005) suggested that the high intrinsic mutation rate of the exon triggered an evolutionary relay between the XLzs and ALEX reading frames. Despite the high rate of nucleotide substitutions within the exon, the number of amino acid replacements is approximately equal between the two reading frames. However, the evolutionary trajectory of the XL exon may be expected to differ in nonprimate mammals. In particular, rodent XL exons are structurally distinct from their primate counterparts: They are longer and lack the internal repeat region primarily responsible for within-primate variation.

To expand our analysis to nonprimate species, we collected sequences orthologous to the human XL exon from all currently available genome assemblies. In addition, we sequenced the exon from a series of rodent species including A. agrarius, A. sylvaticus, M. huberti, P. fumatus, M. caroli, M. hortulanus, M. pahari, and M. spretus. The degree of divergence varied significantly across taxa. For example, the XL exon of human GNAS shows 66% and 83% amino acid identity with those from the cow and rabbit, respectively. Rodent sequences stand out as the most divergent when contrasted with other mammals (60% amino acid identity between mouse and dog and 46% between mouse and cow). In fact, rodent sequences are so divergent from the other mammals that, to perform meaningful analyses, we generated two subalignments: one containing unambiguously aligning mammals and the other containing all rodent species. We further modified alignments by cropping regions showing ambiguities.

The interaction between the products of XLzs and ALEX reading frames imposes a unique set of constraints on the proteins: If XLzs (or ALEX) changes, ALEX (or XLzs) needs to rapidly fix a compensatory substitution to preserve the mutual affinity. Although this cannot be observed directly in our data because such changes are likely to occur within each lineage in rapid succession, the overall effect of this process should be similar rates of amino acid replacements in the two proteins. To test this possibility, we compared nucleotide substitutions between the XLzs and ALEX frames in sequenced species. Classical measures of nucleotide substitution rates for protein-coding regions such as ds and dN (Li 1997) are not directly applicable here because of the interdependence of the two overlapping frames (Krakauer 2000; Pedersen and Jensen 2001; Rogozin et al. 2002). However, these measures can be used in a relative context. Specifically, we estimate nonsynonymous substitution rates separately for each of the two frames (XLdN = ALEXdN) and reestimate the parameters of the model. Likelihood ratio test is then applied to determine whether the two rates are significantly different from each other. Application of this approach to the XL exon of Gnas in nonrodent mammals shows that XLzs and ALEX evolve in a similar manner to their primate counterparts. The ratio of nonsynonymous rates does not differ significantly between the two reading frames (parametric bootstrap P = 0.08) indicating similarity in amino acid replacement rates between XLzs and ALEX (fig. 3). These data are consistent with a scenario in which both proteins evolve under purifying selection (Nekrutenko et al. 2005) and interact with each other, as is the case in humans (Freson et al. 2003).

Surprisingly, the situation is different in rodents where the balance of amino acid replacements is no longer maintained between the XLzs and ALEX reading frames. The ALEX frame accumulates nonsynonymous substitutions significantly faster than the XLzs frame (parametric bootstrap P = 0; fig. 3). Three possibilities may explain the apparent
difference in the evolution of XL exon between rodents and other mammals. First, in rodents, ALEX could be functionally dispensable and therefore devoid of selective constraints. However, the fact that the reading frame-encoding ALEX is preserved in all rodent species considered in the study makes this unlikely. Second, the aggregate biochemical properties of ALEX, a basic protein, may be more important than the sequence itself. In this instance, ALEX would be relatively free to accumulate nonsynonymous changes as long as it stays basic. Here it is difficult to explain why this would uniquely be the case in rodents, however. Finally, rodent ALEX may have a function, which, unlike its mammalian ortholog, does not require interaction with XLzs. This would alleviate binding constraints and therefore the requirement to retain a balance of nonsynonymous substitutions between the two frames. It remains to be seen whether ALEX function is divergent in rodents as described.

Diversity of the Internal Repeat Region within the XL Exon of Gnas

The internal section of the primate XL exon contains imperfect repeated units of variable length translated into amino acid repeats averaging 13 residues in both XLzs and ALEX (Hayward, Kamiya, et al. 1998). The number of repeat units varies across sequenced primates and is polymorphic in humans where deletions and insertions have been reported and implicated in disease (Freson et al. 2003; Gonzalez-Conejero et al. 2004; Nekrutenko et al. 2005). Nonprimate mammals also appear to contain the repeats although they are much less recognizable. This trend is especially pronounced in rodents where repeat units cannot be readily identified by eye. This is true for all examined rodents with the exception of A. agrarius: To our surprise, the sequence of XL exon for that species contained 10 perfect repeat units (10 amino acids each). Apodemus sylvaticus, a sister taxon separated from A. agrarius by approximately 8–10 My (Michaux et al. 2002; Liu et al. 2004), does not contain a single of these 10 amino acid units. Furthermore, between two A. agrarius individuals whose XL exons were sequenced for this study, we observed a polymorphism in the number of repeat units: one of the individuals contains an additional 10 amino acid repeats. To characterize the repeats unambiguously, we applied a de novo protein repeat finder HHrep (Soding et al. 2006) to human (primate), dog (nonprimate, nonrodent), mouse (rodent), and A. agrarius (rodent) sequences (fig. 4). Both nonrodent sequences (human and dog) show a clearly defined area of local self-similarity corresponding to the repeat-containing region. In the mouse, self-similarity regions are less evident and contain many “broken pieces.” Apodemus agrarius also contains multiple small regions of self-similarity, but features a highly ordered region (delineated by a black box in fig. 4) corresponding to agrarius-specific 10 amino acid repeats.

Because the XL exon contains one of three DMRs within the Gnas locus (Coombes et al. 2003), the variation in repeat number might have implication to the methylation status of the gene. Each of the 10 additional repeat units of A. agrarius contains three CpG dinucleotides adding 30 sites that can be potentially methylated and might play a role in the species-specific imprinting status of the locus. Other imprinted genes have also been found to contain similar repetitive sequences in their DMRs, and several studies have suggested a role for repeats in regulating imprinting at these loci (Pearsall et al. 1999; Yoon et al. 2002).

Gs Evolution through Duplication and Emergence of Dual Coding

Comparison of regions homologous between Gnas and Gnal suggests the following sequence of evolutionary
events that gave rise to both genes. First, the 5'-most exon of the ancestral gene duplicated producing exon 1 and XL. The fact that both genes contain XL exons that share sequence homology with the respective first exons (fig. 5) supports this claim. Next, the ancestral gene underwent duplication generating \textit{Gnas} and \textit{Gnal}. Indeed, the two genes have very similar structures with upstream exons XL and 1 being spliced to a series of downstream exons (fig. 1). Our estimate for the time point of this event is between 570 and 528 Ma or after the divergence of vertebrates from invertebrates but before the divergence of tetrapods from fishes (Kumar and Hedges 1998; Blair Hedges and Kumar 2003). \textit{Gnas} and \textit{Gnal} are known to be present in all examined mammals. A recent study documented the presence of the two genes in amphibians (\textit{Xenopus laevis}; Mezler et al. 2001). Genomic assemblies of all sequenced fishes appear to contain both \textit{Gnas} and \textit{Gnal} loci as well (data not shown). In contrast, extensive studies in \textit{Drosophila} suggest the presence of only one Gs gene (Wolfgang et al. 2001).

When did the XL exon of \textit{Gnas} acquire the ALEX reading frame? The terminal amino acids encoded by the XL exon that constitute the \(\beta_1\)-binding domain, consists of four leucine residues, all of which are conserved. The codon positions 2 and 3 of the last leucine residue overlap with the stop codon positions 1 and 2 of the ALEX reading frame. For example, in humans, the codons CTG, CTG, CTT, and CTA (followed by G) encode the four leucine residues. When translated in +1, the last leucine residue overlaps with the stop codon of ALEX (TAG). Leucine can be potentially encoded by six codons (TTA, TTG, CTT, CTC, CTA, and CTG). In mammals, the codon CTG is used most frequently, whereas the codon CTA has the lowest frequency (Nakamura et al. 2000). The last leucine residue of the “THLLLL” motif is encoded by “CTA” in eutherian mammals, all of which show the presence of ALEX reading frame. The use of this suboptimal codon is required to terminate the ALEX reading frame and is therefore under selective constraint. However, the short-tailed opossum \textit{Monodelphis domestica}, a metatherian, lacks the ALEX reading frame and has all four leucine residues encoded by the codon “CTG” (data not shown). A parsimonious explanation of this pattern is that the overlapping reading frame-encoding ALEX first emerged in the eutherian ancestor after diverging from its last common ancestor with metatherians.

\textbf{Discussion}

\textit{Selective Constraints at Gnas/Gnal}

Both \textit{Gnas} and \textit{Gnal} loci are characterized by splice variants that show evidence of monoallelic imprinted expression and differing levels of selective constraint. In particular, the \textit{Nesp55} and XL exons of \textit{Gnas} and the XL exon of \textit{Gnal} show a high rate of interspecies divergence and nonsynonymous nucleotide change, whereas exons 2–13/12 at both loci are largely invariant and conserved. This
correlates with imprinted expression of \textit{Nesp}, \textit{Gnasxl}, and \textit{Gnalx} transcripts, on the one hand, and biallelic expression of the canonical locus transcripts (spliced from exons 1 through 13/12) in the majority of tissues, on the other. At first glance, this pattern appears contradictory. This is because monoallelic expression of a given allele results in its exposure to selection in only half of the individuals carrying it. This would seem to decrease the opportunity for strong directional selection. On the other hand, monoallelic expression carries an important selective penalty. The effectively haploid state of individuals at imprinted loci means that recessive loss-of-function mutations should evoke a deleterious phenotype more often and thereby be efficiently removed by selection. Both these factors should decrease rates of nonsynonymous nucleotide change at the imprinted \textit{Nesp}55 and XL exons. The conflict theory, on the other hand, may explain the contrary observation.

**Conflict Theory and \textit{Gnas/Gnal}**

The conflict or kinship theory (Moore and Haig 1991) for the evolution of genomic imprinting is an intraindividual extension of parent–offspring conflict theory. A mother is selected to maximize her lifetime reproductive success by spreading her costly parental investment between offspring and litters. This means that, under many circumstances, offspring are selected to extract more resources than a mother is selected to provide (Trivers 1974). We expect this to be the case for both maternally and paternally derived alleles in offspring. But we also expect this process to be limited, in a given offspring, by kin selection acting on gene copies in littermates or in future siblings. In family groups, matrilineal and patrilineal kin are not always symmetrically related. So multiple paternity, within or between litters, can lead to a situation in which paternal alleles are relatively less constrained by kin selection than maternal alleles. When we allow alleles to show levels of expression that depend on their parental origin, we expect an arms race to develop in which increases in paternal expression of resource-demanding genes are balanced by decreases in maternal expression of the same. The reverse dynamic is expected for resource-conserving genes, and these responses can occur in the same or in different alleles across generations leading to complete or polymorphic imprinting, respectively. Just this arms race scenario was first offered as an explanation for the evolution of imprinting at two loci located on different chromosomes in mice (Haig and Graham 1991). \textit{Igf2}, on chromosome 7, is a growth factor gene, paternally transmitted null mutations in which lead to severe prenatal growth retardation (DeChiara et al. 1990, 1991), whereas mutations affecting the maternally expressed growth suppressor, \textit{Igf2r}, on chromosome 17 result in growth retardation (Barlow et al. 1991). \textit{Igf2r} is a receptor that sequesters \textit{Igf2} and targets it for intracellular degradation (Braulke et al. 1999). Work on a variety of imprinted genes support the key prediction of conflict theory—paternal genes tend to promote fetal growth (and therefore prenatal resource demand), whereas maternal genes suppress it (Haig 2004). In general, maternal and paternal genes are expected to be in conflict in respect of any offspring phenotypes with effects on parental investment or, more generally still, conflict is expected for any phenotypes affecting altruistic interactions between asymmetrically related kin.

Returning to the \textit{Gnas} and \textit{Gnal} loci, we can see that exons spliced to imprinted transcripts may be expected to participate in arms races of the sort described. Although neither prenatal nor immediate postnatal growth effects have been described at either locus, XL\textit{zxs} (from \textit{Gnasx}) is known to be a critical regulator of postnatal energy metabolism and suckling (Plagge et al. 2004), the latter function aligning well with predictions from conflict theory. \textit{Nesp}55 is also known to affect reactivity to novel environments, a phenotype that might influence dispersal from the natal area (Plagge et al. 2005) and that might be subject to selection under conflict (Isles et al. 2002). So does participation in an arms race explain high rates of sequence divergence in the \textit{Nesp} and XL exons? Perhaps, but maybe only if gene function itself, not just allelic expression, is subject to antagonistic coevolution. One obvious means by which this might be realized is by the evolution of defective, inhibitory forms of G-proteins. But mouse and human XL\textit{zxs} are capable of signal transduction in the same manner as \textit{Gaz} (Basteppe et al. 2002; Linglart et al. 2006), and \textit{Nesp}55/\textit{NESP}55 have unrelated molecular functions. The significance of the \textit{Nesp} alternative reading frame and its gene product is rather more promising and is discussed below. Another possibility is that the regions examined are subject to high rates of sequence divergence precisely because they are involved in the regulation of imprinting. In this regard, the presence of functionally significant DMRs has already been noted.

Here, though, we must sound a note of caution. Simple arms races of the sort described might occur in the evolutionary “blink of an eye” leaving little evidence of interspecies divergence of the kind seen. Although differences in life histories or mating systems may account for changing selective pressures under conflict, this is not a sufficient explanation, and we need to explain why selective pressures should be so particularly diverse in this instance compared with others. So what are we left with by way of explanation?

**Imprinting and an Escape from Pleiotropy**

Given the costs of monoallelic expression described earlier, the existence of imprinting at the \textit{Gnas} and \textit{Gnal} loci is surprising. This is because both genes encode crucial cell signaling functions with the former showing near ubiquitous expression. Indeed, heterozygous inactivating mutations throughout the human \textit{GNAS} gene have severe phenotypic consequences (Aldred and Trembath 2000; Weinstein et al. 2004). On the other hand, for genes with pleiotropic effects, it is more likely that at least one of these will be subject to selection under conflict theory. Further, there are ways to minimize the impact of deleterious recessives and the \textit{Gnas} and \textit{Gnal} loci illustrate two of these:

1. **Tissue-specific imprinting**: the canonical \textit{Gnas} transcript is subject to monoallelic expression in a subset of tissues (Yu et al. 1998; Hayward et al. 2001; Germain-Lee et al. 2002; Mantovani et al. 2002; Liu et al. 2003). The impact of deleterious recessive alleles would be
reduced by the presence of wild-type alleles in most tissues because they have biallelic expression.

2. Alternative splicing: alternative transcripts are able to exert phenotypic effects in a manner that is at least partially independent (perhaps fully independent for Nesp; Plagge et al. 2005) of the canonical locus transcript.

These two explanations are essentially combined in the case of the Nesp and Gnasxl transcripts that are expressed in only a subset of tissues and encode evolutionarily novel functions. Pleiotropic effects of mutations at the Nesp and XL exons during evolution might therefore be fewer or less severe than those associated with mutations in the canonical or shared downstream exons.

So we argue that high rates of sequence evolution at the Nesp and XL exons are not surprising because changes in function at these exons are subject to fewer constraints imposed by correlated, independent phenotypes associated with the Gnas and Gnal canonical transcripts. Taken together, with the idea that novel selection pressures are introduced by imprinting, we believe this may explain the observed higher rate of change at these exons contrasted with their more ancient and functionally entrenched neighbors.

Is There a Role for the Nesp Alternative Reading Frame?

We describe conservation of an alternative reading frame within the Nesp exon, the start codon for which overlaps with the Nesp55 stop codon. The function of this open reading frame (ORF) is difficult to assess without in vivo evidence, but it seems likely from the evidence we present to encode a protein with effects visible to selection as determined by the transition/transversion ratio mutations across codon positions. This is interesting because genetic evidence suggests that the Gnas locus is characterized by conflicting maternal and paternal gene functions.

Maternally versus paternally transmitted null mutations affecting exon 2 of Gnas affect postnatal energy metabolism in mice in opposing ways (Yu et al. 1998, 2000). Two-day-old Gnas exon 2 m−/+ pups, for example, showed enhanced lipid accumulation in adipose tissues compared with wild-type littermates, whereas Gnas exon 2 +/p− pups showed decreased lipid accumulation (Yu et al. 2000). Underlying these observations were changes in the expression levels of Ucp1 in brown adipose tissue, a gene required for activation of nonshivering thermogenesis in this tissue (Yu et al. 2000).

Several mutations affecting other Gnas locus exons have been characterized with broadly opposite postnatal effects on paternal versus maternal inheritance. For example, a point mutation in exon 6 of Gnas (the Oed−Sml mutation; Skinner et al. 2002) resulted in growth retardation upon paternal inheritance (Sml) and a gross edema upon maternal inheritance (Oed; Cattanach et al. 2000; the latter effect was also noted in Gnas exon 2 m−/+ pups; Yu et al. 1998). A partial phenocopy of the Gnas exon 2 m−/+ phenotype was also obtained upon maternal inheritance of a null mutation in exon 1 of Gnas (Chen et al. 2005; Germain-Lee et al. 2005). But paternal inheritance of the same mutation (Gnas exon 1 +/p−) did not result in growth retardation or impaired survival before weaning (Chen et al. 2005). These effects were instead noted in pups with paternal inheritance of a null mutation in the Gnasxl exon, which also showed impaired suckling and hypoglycemia thereby suggesting a critical role for XLzs in successful feeding and in regulating postnatal energy metabolism (Plagge et al. 2004).

We may therefore infer that opposite effects are caused by maternally expressed transcripts possessing exons 1, 2, and 6, on the one hand, and paternal transcripts possessing the XL exon and exons 2 and 6 of Gnas, on the other. A role for ALEX in paternal loss-of-function phenotypes, or for transcripts truncated after exon 3 of Gnas (see Crawford et al. 1993; Pasolli et al. 2000; Plagge et al. 2004), is apparently discounted by growth retardation in Sml pups. However, roles for these paternal factors in respect of subtle or unaccounted differences between paternal loss phenotypes cannot be discounted, and their significance remains to be established by targeted mutagenesis experiments.

The identity of the proteins involved in the maternal loss-of-function phenotypes is likewise not certain, but a role for Nesp55 is highly unlikely given that a mutation affecting the Nesp reading frame has no discernable effects in neonates (Plagge et al. 2005). That the Nesp alternative reading frame is involved seems unlikely because of the overlap between the Gnas exon 1 m−/+ and Gnas exon 2 m−/+ phenotypes alluded to above. We are therefore left with the likely answer that loss of the canonical Gnas transcript in tissues with imprinted (maternal) expression is responsible for the phenotypes described. But again a role for the Nesp alternative reading frame cannot be entirely discounted.

Some differences have been noted between Gnas exon 1 m−/+ and Gnas exon 2 m−/+ pups, the latter showing improved survival, higher body weight, and lacking neurological defects seen in the former (Chen et al. 2005). Further, at adult stages, the two genotypes appear to show opposite phenotypes: Gnas exon 1 m−/+ mice are insulin resistant (Chen et al. 2005), whereas Gnas exon 2 m−/+ mice are insulin sensitive (Yu et al. 2001). Perhaps, a truncated protein produced by the Nesp alternative reading frame can interfere with insulin signaling. Gzs subunits lacking the binding sites for beta and gamma subunits might competitively interact with activated G-protein–linked receptors reducing signal transduction through functional Gzs subunits. We would therefore welcome further investigation of the in vivo significance of the Nesp alternative reading frame.

Coevolution within the XL Exon

From the perspective of conflict theory, the existence of the ALEX reading frame within the Gnasxl exon is difficult to explain. Naively, we might imagine that a negative interaction between XLzs and ALEX proteins (see Introduction; Freson et al. 2001, 2003) implies that an evolutionary conflict exists within the paternal genome. But conflict theory can only account for conflict arising from asymmetries of relatedness between different gene copies. Perhaps, we might suppose that trans effects originating from the maternal allele are responsible for conflicting functions, but how then can we account for the maintenance of the ALEX
reading frame? Imprinting, under conflict, is essentially a conditional strategy in which alleles optimize their expression depending on circumstance. We would not therefore expect a gene to maintain a reading frame that is deleterious to it on paternal inheritance. It therefore seems unlikely that any explanation for the evolution of the ALEX reading frame, which probably emerged after the divergence of eutherians from metatherians, will focus on evolutionary conflict as envisaged in the conflict theory.

We analyzed nonsynonymous substitution rates in the XL and ALEX reading frames in rodents finding rates in the latter to be higher than to findings in other mammals (and our previous finding in primates) where rates were not significantly different. We believe the most likely explanation for this is that rodent ALEX has acquired a function that does NOT involve binding to XLzs thereby alleviating the requirement for complementary substitutions hypothesized in nonrodents. It is possible that the basic amino acids in ALEX lead to relatively promiscuous binding and that XLzs/ALEX binding in rodents is abrogated by the absence of repeats. We discovered that imperfect tandem repeats identified in nonrodents were absent in all the rodents we studied with the exception of A. agrarius (which further displayed a repeat polymorphism between two individuals). It is again difficult to account for such differences in terms of conflict theory because both A. agrarius and the closely related A. sylvaticus, which lacked repeats, show evidence of multiple paternity in the wild (Baker et al. 1999). We therefore have no reason to suppose that changes in family structure lie behind changes in the GnasXL exon. There is clearly an interesting evolutionary dynamic in respect of the interaction between XLzs and ALEX, and we would welcome more work on the functions of the ALEX protein in mouse and in human, which may be expected to differ from these data.

Conclusion

The Gnas and Gnal paralogous genes are members of the Gx family of genes that encode the alpha subunits of G-proteins involved in many key signal transduction pathways. Both genes are characterized by the use of alternate upstream promoters and first exons and show evidence of imprinting. We have analyzed the pattern of substitution rates across both loci revealing an increased \( \frac{dN}{dS} \) ratio in the partition containing the alternate start exons. This likely correlates with the relatively less constrained functions we believe are associated with alternate imprinted transcripts versus the entrenched functions associated with the housekeeping roles of Gnas and Gnal canonical transcripts. But it may also or only be a consequence of enhanced mutation rates associated with CpG methylation at DMRs at both loci.

Considering the species distribution of the loci and their components further leads us to infer that the emergence of the extra long XL exons at both loci preceded the divergence between them, the latter event occurring \( \sim 570 \) Ma. To some degree, therefore, the genetic complexity of these loci likely preceded the emergence of imprinting in mammalian lineages meaning that imprinted expression at both loci was independently acquired. Perhaps, one reason for the regulatory complexity associated with imprinted loci in general is that genes with pleiotropic functions tend to be not only those that attract imprinting under conflict but also those that impose the most constraint on imprinting. As argued above, it then seems likely that evolutionary change outside the canonical transcript and its exons is favored.

We also analyzed transition/transversion ratios in the Nesp alternative reading frame and nonsynonymous substitution rate ratios between the two reading frames in the XL exon (the Gsalpha and ALEX ORFs) in a wide range of mammals. In the former, we provide strong evidence for the functional importance of the putative truncated Gzs produced from the Nesp transcript, and we go on to speculate whether this transcript may play a role in the opposite postnatal phenotypes observed in mice. A possible role for such a protein may be to negatively regulate the response to insulin signaling although because this is based on comparisons between different experiments and not on direct experimental data, this cannot be asserted with any force. It also appears that the ALEX protein product, the reading frame for which is maintained in all eutherians examined to date, has an important function.

Previously, we have shown that nonsynonymous substitution rates in both the Gnas and ALEX overlapping reading frames were balanced in the primate phylogenetic tree in a manner suggesting coevolution between the two reading frames (Nekruttenko et al. 2005). This correlated with human data suggesting an interaction between the two proteins (Freson et al. 2001, 2003). Here we extend this analysis to other mammals and show that, although this remains the case for nonrodent mammals, in the Muridae, in particular, there appears to be a higher rate of nonsynonymous change in the ALEX reading frame. We believe this is suggestive of divergent ALEX function in rodents, possibly involving interactions between ALEX and a third protein. A particularly high rate of overall sequence divergence and the presence and absence of tandem repeats within this exon in closely related rodent species also suggests that ALEX function may be rapidly diverging within this lineage for reasons we do not fully understand. Clearly, much remains to be understood about this enigmatic corner of the immensely complex Gnas locus.

Supplementary Material


Acknowledgments

Authors are grateful to Gavin Kelsey for his advice and to Kateryna Makova and Paula Goetting-Minesky for experimental support. Dr Robert J. Baker provided rodent DNA samples for this study. Special thanks to Wen-Yu Chung, Radek Szklarczyk, and Sergei Kosakovsky-Pond.
Clues to the Evolution of the Gnas and Gnal Loci 2755

for analysis ideas. This study was supported by the Beckman Young Investigator Award to Anton Nekrutenko and by funds from Penn State University. The authors declare that there is no conflict of interest.

Literature Cited


Takashi Gojobori, Associate Editor

Accepted October 1, 2008