Using Gene-History and Expression Analyses to Assess the Involvement of LGI Genes in Human Disorders

Wenli Gu,*† Yann Gibert,† Thierry Wirth,† Andrea Elischer,‡ Wilhelm Bloch,‡ Axel Meyer,† Ortrud K. Steinlein,* and Gerrit Begemann†

*Institute of Human Genetics, Ludwig Maximilians University Munich, University Hospital, Munich, Germany; †Lehrstuhl Zoologie/Evolutionsbiologie, Department of Biology, University of Konstanz, Konstanz, Germany; and ‡Department of Molecular and Cellular Sport Medicine, German Sport University Cologne, Cologne, Germany

Mutations in the leucine-rich, glioma-inactivated 1 gene, LGI1, cause autosomal-dominant lateral temporal lobe epilepsy via unknown mechanisms. LGI1 belongs to a subfamily of leucine-rich repeat genes comprising four members (LGI1–LGI4) in mammals. In this study, both comparative developmental as well as molecular evolutionary methods were applied to investigate the evolution of the LGI1 gene family and, subsequently, of the functional importance of its different gene members. Our phylogenetic studies suggest that LGI1 genes evolved early in the vertebrate lineage. Genetic and expression analyses of all five zebrafish lgi genes revealed duplications of lgi1 and lgi2, each resulting in two paralogous gene copies with mostly nonoverlapping expression patterns. Furthermore, all vertebrate LGI1 orthologs experience high levels of purifying selection that argue for an essential role of this gene in neural development or function. The approach of combining expression and selection data used here exemplarily demonstrates that in poorly characterized gene families a framework of evolutionary and expression analyses can identify those genes that are functionally most important and are therefore prime candidates for human disorders.

Introduction

The final annotation of the human genome has identified many new gene families. When one member of a gene family is identified to be related to a human disease, other members of the gene family are often considered as candidate genes for similar disorders. However, the evaluation of each single gene is often both costly and time consuming. The progress in different genome databases offers the possibility to identify orthologs of human genes in a number of other organisms and to study the evolution of these genes. The zebrafish is increasingly used to model human development and disease (Zon 1999; Dooley and Zon 2000). The physiological functions and expression patterns of many orthologous genes between zebrafish and humans have been conserved to various degrees such that mutants and knockdowns of the zebrafish orthologs of human disease genes have established models for a wide spectrum of human phenotypes (Zon and Peterson 2005).

A number of mutations in the human LGI1 gene have been shown to cause autosomal-dominant lateral temporal lobe epilepsy (ADLTE), a rare idiopathic epilepsy (Gu, Brodkorb, and Steinlein 2002; Kalachikov et al. 2002; Morante-Redolat et al. 2002). Idiopathic epilepsies are those in which a symptomatic background is neither detected nor suspected, but a genetic etiology is likely or proven. Most idiopathic epilepsies are caused by ion channel mutations, implicating an etiology based on imbalances in synaptic transmission or neuronal excitability (Steinlein 2004). Surprisingly, LGI1 does not appear to encode an ion channel (Kalachikov et al. 2002), thus the disease mechanisms of the LGI1 mutations remain unknown and may open a new aspect of epilepsy pathogenesis. Additionally, LGI1 is considered as a possible new member of the emerging subfamily of tumor suppressor genes referred to as “metastasis suppressors” (Kunapuli et al. 2004); a number of glioma cell lines and malignant brain tumors show a strong reduction of LGI1 expression (Chernova, Somerville, and Cowell 1998; Krex et al. 2002; Besleaga et al. 2003), while, conversely, forced expression of LGI1 in glioma cells lacking endogenous LGI1 expression inhibits their proliferation and invasiveness (Kunapuli et al. 2004).

Previously we cloned three additional members of the human LGI gene family, LGI2–4 (Gu et al. 2002). The human LGI1 proteins share 65%–75% sequence identity with each other, and all contain 4.5 leucine-rich repeats (LRR) in the N-terminal part and seven epitempin (EPTP) repeats in the C-terminal part. LRRs have been suggested to participate in protein-protein interactions (Kajava 1998; Kobe and Kajava 2001). The EPTP repeats were identified in only two other genes, including MASS1/VLGR1, which is mutated in a mouse model for epilepsy (Skradski et al. 2001; Gibert et al. 2005). The genomic localizations of human LGI2–4 overlap with candidate regions for several other epilepsy syndromes and malignancies, LGI2–4 therefore being considered as candidate genes for these disorders.

Several studies on the evolutionary pressures acting on disease-related genes have equivocally suggested that purifying selection is indicative of essential (disease-related) genes (e.g., Yang, Gu, and Li 2003). Using the leucine-rich, glioma-inactivated (LGI) gene family as a model, we tested the usefulness of an integrated framework of evolutionary and expression analyses to make a prediction of which LGI gene members are most likely related to human disorders and which should therefore be given preference in candidate gene evaluation. We screened sequence databases of different organisms for previously undiscovered LGI orthologs and analyzed the expression of all five lgi genes in zebrafish embryos and adult brains. Moreover, we compared the expression patterns and genomic localizations to study the evolutionary history and determined the force and type of natural selection acting on the LGI gene family.
Materials and Methods
Fish Stocks, Sequence Data and Phylogenetic Analyses, Mapping and Syntenic Analyses

The data are available in the Supplementary Materials and Methods section.

In Situ Hybridization and Photography

Whole-mount in situ hybridization of zebrafish embryos were performed as previously described (Begemann et al. 2002). To prevent melanization in larvae older than 30 hours post fertilization (hpf), embryos were exposed to 0.2 mM 1-phenyl-2-thiourea. Embryos were mounted in 70% glycerol and examined with a Zeiss Axiophot microscope. Images were processed using Zeiss Axiovision and Adobe Photoshop software.

Results
Identification of Nonhuman LGI Genes and Cloning of Zebrafish Orthologs

To search for orthologs of the human LGI genes in other species, we performed Blast searches in different species whose genomes are fully or partially available. Whereas orthologs were identified in chimpanzee (Pan troglodytes), chicken (Gallus gallus), zebrafish (Danio rerio), and puffer fish (Takifugu rubripes, Tetraodon nigroviridis) genomes, no LGI orthologs could be identified from the invertebrate genomes of nematode (Caenorhabditis elegans), the fruitfly (Drosophila melanogaster), and the ascidian Ciona intestinalis (Table S1, Supplementary Material online). With the exception of the puffer fish genes, all putative LGI homologs were also identified in expressed sequence tag (EST) databases and hence can be considered to be transcribed in vivo.

Phylogeny of the LGI Gene Family

Based on the alignments of all retrieved genes, we constructed a phylogeny of the LGI gene family (fig. 1). In the absence of LGI sequence data from suitable non-vertebrate out-groups and due to the lack of related vertebrate genes with considerable sequence similarity, the tree is unrooted. Tree topologies for nucleotides in first and second codon positions and amino acids are identical and allow an unambiguous assignment of orthologous relationships between fish and mammalian genes. The tree
Syntenic Relationships Between Zebrafish and Human LGI Genes

All five zebrafish *lg* genes map to different chromosomes, suggesting that none of them arose by tandem duplication (table 1). Based upon the mapped genes surrounding both zebrafish and human *lg* genes, we determined whether the human and fish *LGI* loci exhibit conserved synteny (fig. 2). The zebrafish *lgi1a* and *lgi1b* genes map to chromosomes 13 and 12, respectively, which have been shown to share other paralogous gene pairs, including the annexins *anxa1a/b* (Farber et al. 2003) and paired box genes *pax2a/b* (Woods et al. 2000). The human ortholog of these genes maps to 10q23–24, and we found conserved syntenies between zebrafish *lgi1b* and human *LGI1* at the level of local gene order. Within a region of approximately 160 kb both *lgi1b* and *LGI1* are flanked by genes for *phosphodiesterase 6C* (PDE6C) and *retnin binding protein 4* (RBP4). The putative orthologs of several genes like the *early growth response gene 2* (ERG2) or the *fibroblast growth factor gene 8* (FGF8) flanking human *LGI1* more distally were found at greater distances from *lgi1b* and *lgi1a*, respectively. The lack of supercontigs containing *lgi1a* presently precludes a local synteny analysis of flanking genes. Taken together, the phylogeny and syntenic relationships of the *LGI1* orthologs strongly suggest that zebrafish *lgi1a* and *lgi1b* are paralogs that arose during duplication events involving larger chromosomal regions.

Zebrafish *lgi2a* and *lgi2b* map to chromosomes 9 and 1, respectively, which also harbor paralogous genes of *en-grailed* (*eng1a, eng1b*) and *distal-less homeobox* (dlx2a, dlx2b) (Taylor et al. 2003). Moreover, we identified several genes close to *LGI2* on human chromosome 4 with putative orthologs on zebrafish chromosome 1 (fig. 2), including *superoxide dismutase 3* (SOD3) and *cholecystokinin type A receptor* (CCK-AR). Together with the phylogenetic topology of the gene tree (fig. 1C), these data establish that *lgi2a/b* are paralogs.

Finally, human and zebrafish *LGI3* genes map to human and zebrafish chromosomes 8, together with orthologs of four other genes (fig. 2). Among them is the *SRC-like-adaptor gene* (SLA), which has a putative ortholog (*sla*), that is located within 60 kb of *lgi3*. We were unable

### Table 1

<table>
<thead>
<tr>
<th>Gene</th>
<th>Closest Flanking Marker</th>
<th>Chromosome (RH panel)</th>
<th>Ensembl Map Position (Zv4-scaffold Number)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>lgi1a</em></td>
<td>3.67 cR from Z5643</td>
<td>Chr. 13 (LN54)</td>
<td>N/A (NA16905, NA9673)</td>
</tr>
<tr>
<td><em>lgi1b</em></td>
<td>3 cR from Z6920</td>
<td>Chr. 12 (T51)</td>
<td>Chr. 12 (1018)</td>
</tr>
<tr>
<td><em>lgi2a</em></td>
<td>Not very strong mapping</td>
<td>N/A</td>
<td>Chr. 9 (825)</td>
</tr>
<tr>
<td><em>lgi2b</em></td>
<td>20.70 cR from Z23059</td>
<td>Chr. 1 (LN54)</td>
<td>Chr. 2 (124.14)</td>
</tr>
<tr>
<td><em>lgi3</em></td>
<td>Not very strong mapping</td>
<td>N/A</td>
<td>Chr. 8 (756)</td>
</tr>
</tbody>
</table>

*Zv4-scaffold124.14 is wrongly placed on chromosome 2, as the BAC ends of this region align with clones mapped in ctg85 in chromosome 1 (M. Caccamo, Wellcome Trust Sanger Institute); Abbreviations: N/A, not available; BAC, bacterial artificial chromosome; and Chr., chromosome.*
to find syntenic clusters surrounding the LGI3 loci within a range of up to 1 Mb. This suggests that the gene orders on these chromosomes have been extensively rearranged since the split between mammals and teleosts.

Zebrafish lgi Gene Expression Patterns During Development and in Adult Brain

We examined the embryonic expression patterns of all zebrafish lgi genes by whole-mount in situ hybridization. Expression of lgi1a is first evident in the ventral diencephalon and at 24 hpf strong expression is observed in the developing eyes, in the ventral midbrain and hindbrain, and in the peripheral spinal cord (fig. 3A–D). By 48 hpf lgi1a is strongly expressed in the retinal ganglion cell layer, the diencephalon, and along the ventral aspect of the hindbrain (fig. 3E–H). Notably, all lgi1a expression domains are in neural tissues. lgi1b is expressed at 24 hpf in presumptive telencephalic and diencephalic bands and cranial paraxial mesenchyme. At 48 hpf, lgi1b transcripts are detected in the optic tectum, the cerebellum, and in the zone of migratory neurons that originated in the rhombic lip. Expression is further observed in the dorsal thalamus and in the retinal ganglion cell layers (fig. 3I–L). Overall, lgi1a expression is predominant in ventral parts of the mid- and hindbrain, while lgi1b is more dorsally restricted in this region. In situ polymerase chain reactions (PCRs) on adult transversal brain sections (fig. 3M and N) show that lgi1a and lgi1b are expressed in the outer layer of the periventricular gray zone (pgz) of the optic tectum, an area rich in tectal neurons. lgi1b, in addition, is strongly expressed in the cerebellum. Both genes colocalize with nuclear areas of ganglion cells. At this level of resolution we could not detect expression in adult brain glial cells. In contrast, expression of both lgi2 paralogs is generally restricted to a few cells of putative endodermal origin during embryogenesis. Both genes are expressed in trigeminal ganglion cells and in a few cells in the posterior head (fig. 4A–C). More prominently, lgi2a is transiently detectable in dorsal spinal cord neurons. Finally, lgi3 is expressed in cranial mesodermal cells and in a few cells on each side of the otic vesicle (fig. 4D and not shown). lgi3 appears to be coexpressed with lgi1a in the peripheral spinal cord in 1- and 2-day-old embryos and is detected in a reiterated symmetrical pattern of cells in the ventral hindbrain (fig. 4E and F).

Different Types of Selection Among Family Lineages of LGI Genes

To test for possible differences in evolution rates after the duplication events or during the course of subfunctionalisation, we first tested for the possibility that the data set has already lost phylogenetic information due to
accumulation of mutations and the resulting saturation. Plotting of transition and transversion rates as a function of genetic distances suggested that transitions have reached saturation (fig. 1A).

We therefore applied additional statistics in order to measure saturation substitution at first, second, and third codon positions separately using the Xia index (Xia et al. 2003). This index allows us to judge whether a set of aligned sequences is useful in phylogenetics or not. The index of substitution saturation is defined as $I_{SS} = H/H_{FSS}$. When $I_{SS}$ approaches 1, the sequences experienced severe saturation. However, this is only useful in theory because phylogenetic reconstructions will fail to recover the true tree long before the full substitution saturation is reached. Therefore, another parameter $I_{SS,C}$ has to be computed at which the sequences will begin to fail to recover the true phylogeny. Once $I_{SS,C}$ is known for a set of data, we can infer the $I_{SS}$ value from the sequences and compare it to $I_{SS,C}$. If $I_{SS}$ is not smaller than $I_{SS,C}$, we can conclude that saturation will interfere with phylogenetic analyses. For the third codon position of the LGI coding sequences, the observed $I_{SS}$ value of 0.913 is significantly larger than the $I_{SS,C}$ value of 0.723 (95% confidence interval, 0.844 < $I_{SS}$ < 0.981). Thus, $I_{SS} > I_{SS,C}$ and the third bases are of limited value for phylogenetic reconstruction. In contrast, first and second codon positions showed an $I_{SS}$ value of 0.770 that is significantly larger than the $I_{SS}$ value of 0.582, which confirmed that there is little saturation at these sites, indicating that reliable phylogenetic signal is contained in the first two codon positions. Also, maximum likelihood (ML) mapping confirmed that there is a sufficient amount of phylogenetic information, with 87.7% fully resolved quartets at third base and 94.8% fully resolved quartets at first and second bases (fig. 1B).

We next estimated the likelihood of the data under a unique $\omega$ ratio among all lineages. The log-likelihood under this model was $l_0 = -20,508.56$, with parameter estimates $k = 1.50$ and $\omega = 0.121$ (Table S2, Supplementary Material online). This $\omega$ ratio was an average over all sites and lineages. In a second step we tested if more complex models (with different selection pressure) among the LGI orthologs versus the other groups of LGI orthologs are more likely (see Supplementary Materials and Methods). This was in fact the case and the likelihood value under the H1 model was $l_1 = -20,449.06$. Comparison of the $2\Delta l = 2(l_1 - l_0) = 2 \times 59.5 = 119$ with the $X^2_{df} = 6.63$ suggests rejection of the one ratio model. The partitioning of the selection pressure into four categories, one for each LGI gene (fig. 1C), was the model which best fit the data (Table S2, Supplementary Material online). Estimates of the $\omega$ ratios (Table S2, Supplementary Material online) determined that the selection pressure differs among the four LGI genes. LGI1 and LGI4 are under very strong negative selection, whereas the LGI2 and LGI3 genes, although being under purifying selection, seem to be under more relaxed selection pressure.

### Variation in Selective Pressure Across Codon Sites

Parameter estimates and log-likelihood values under models of variable $\omega$ among sites are presented in Table S3 (Supplementary Material online). Model M0 poorly fits the data when compared to model M3. The latter model involves four more parameters than M0, and the likelihood ratio test (LRT) statistic $2\Delta l = 1.061.66$ is much greater than the critical $X^2_{df} = 13.28$ with df = 4. The results suggest variation in selective pressure among amino acid sites. Moreover, all three models that allow for the presence of sites under selection, i.e., M2 (selection), M3 (discrete), and M8 ($\beta$ and $\omega$) better fit the data than alternative models that do not allow for selection (Table S3, Supplementary Material online). A striking feature under the “selection” models is that all sites seem to be under purifying selection, and no single site under positive selection was detected. Posterior probabilities for site classes calculated under M3 (discrete) are plotted in Figure S1 (Supplementary Material online). Six out of 10 amino acids mutated in human ADLTE exhibit high selection pressure, an observation which is in agreement with the role these mutations are assumed to play in the pathogenesis of this rare epilepsy.

ML estimation suggests that the three site classes are in proportions $P_0 = 0.334, P_1 = 0.479$, and $P_2 = 0.188$, with the ratios $\omega_0 = 0.016, \omega_1 = 0.129$, and $\omega_2 = 0.399$, respectively. (Table S3, Supplementary Material online). Those proportions correspond to the prior probabilities that any site belongs to each of the three classes. For example, the posterior probabilities for site 5 (L) are 0.000, 0.006, and 0.994, and this site is therefore under purifying selection, though belonging to the lower constraint class. The probabilities for site 42 (C) are 0.990, 0.001, and 0.000, showing that this position is extremely constrained and under very strong purifying selection ($\omega = 0.016$). The results obtained from models M2 (selection) and M8 ($\beta$ and $\omega$) were similar (data not presented). The only clear pattern obtained from the posterior probabilities for site classes with different selection pressures for amino acids sites along the LGI sequences is a 40-aa-long stretch under moderate negative selection at the N-termini. The rest of the molecule seems to be more constrained (Fig. S1, Supplementary Material online).

### Discussion

#### Evolution of the LGI Gene Family

Our analyses demonstrate that orthologs of the LGI gene family are absent from invertebrate genomes, as far as their sequences are currently available, and therefore suggest that the LGI gene family originated in the evolutionary lineage leading to the vertebrates. Our finding that all zebrafish lgi genes are predominantly expressed in tissues of neural origin suggests that this gene family may have been involved in the evolution of the vertebrate brain. Phylogenetic relationships and topology of the four mammalian LGI family members (fig. 1C) indicate an origin of the gene family through two rounds of gene or genome duplications. In this scenario, each of the two gene pairs LGI1/LGI4 and LGI2/LGI3 had one ancestral precursor gene. These two ancestral genes themselves may have arisen from a common “proto-LGI” gene. The fact that mammalian genomes have evolved by a diversity of duplication events, which probably included two complete genome duplications early during vertebrate evolution (Lynch and Conery
2000; Wang and Gu 2000; Wolfe 2001; Samonte and Eichler 2002; Jaillen et al. 2004), supports this interpretation of LGI gene family evolution. Irrespective of the mechanism, we predict that a single LGI homolog is present at the root of the vertebrate lineage, the ortholog of which may await identification in urochordates or cephalochordates (e.g., Amphioxus).

In actinopterygians (ray-finned fish), which have undergone an additional genome duplication (Amores et al. 1998; Taylor et al. 2003; Jaillen et al. 2004; Postlethwait et al. 2004; Vandepoele et al. 2004), two pairs of paralogous lgi1a/b and lgi2a/b genes are found. We were able to establish the orthologous relationships between the four mammalian and five zebrafish LGI genes, which suggest duplications of LGI1 and LGI2 genes. The loss of one copy of lgi3 has to be postulated if the duplication of LGI genes is indeed due to the additional genome duplication in actinopterygians. The branch lengths of the fish lgi2 genes are larger than those of mammals, which we interpret as a sign of accelerated rates of evolution within this subfamily, and particularly for lgi2b. Because this gene has been lost in the lineage leading to the puffer fish, it might have been functionally redundant after the duplication event. Its persistence in zebrafish thus suggests that Lgi2b may have acquired a novel function in zebrafish.

LGI4 appears to be absent from zebrafish and puffer fish. The most probable scenario is that LGI4 was lost in the lineage leading to the ray-finned fish. In the human and mouse genomes, LGI4 is flanked by two FXYD domain containing ion transport regulator genes, FXYD1 and FXYD3, at the 5' and 3' ends, respectively. Interestingly, the putative zebrafish ortholog of FXYD1 (fi25c12) maps to chromosome 15 (Zv4_scaffold1327.1), while the fish ortholog of FXD3 is present on chromosome 16. Thus the absence of LGI4 orthologs in the zebrafish and puffer fish may be explained by a high degree of genome rearrangements entailing degeneration or deletion of the LGI4 locus since the split of the ray-finned and lobe-finned fish lineages. Alternatively, LGI4 may have originated from a duplication of LGI1 in the lineage leading to the sarcopterygians (lobe-finned fish) and also the mammals. Unfortunately, it is not possible to date duplication events within the LGI family because third codon positions have reached saturation and remaining codon positions are under selection pressure.

Expression of Duplicated Zebrafish LGI Genes
Suggests Subfunctionalization

The knowledge of embryonic gene expression patterns can shed light on the developmental processes linked to LGI gene activity. The two zebrafish LGI1 orthologs are expressed in partly complementary patterns. For example, lgi1a and lgi1b are expressed in nonoverlapping domains in ventral and dorsal parts of the fore-, mid-, and hindbrain, respectively (fig. 3). This finding suggests partitioning of the original regulatory elements, followed by subsequent degenerative changes in both duplicates. This model of subfunctionalisation after duplication is known as the Duplication-Degeneration-Complementation model (Force et al. 1999), in which the combined expression patterns of the paralogous genes constitute the expression pattern of the original. The lgi1 paralogs also share common sites of gene expression, indicating that they may act in a redundant fashion in these areas. Similar to the situation in the mouse brain (Kalachikov et al. 2002) zebrafish lgi1 gene expression in the adult brain is associated with dense packings of neurons (fig. 3O and P), while evidence for glial expression could not be found.

Expression of the remaining mammalian LGI genes had so far only been studied by semiquantitative PCR methods in adult mice (Nagase, Kikuno, and Ohara 2001; Gu et al. 2002; Runkel, Michels, and Franz 2003). Zebrafish lgi2a and lgi2b transcripts are restricted to a few cells only with coexpression being restricted to the trigeminal ganglia. Moreover, they are predominantly, if not exclusively, expressed in neural tissues. lgi3 appears to be coexpressed with lgi1a in the spinal cord and is expressed in the ventral hindbrain, although in a different pattern than lgi1a. Remarkably, lgi3 is expressed in the developing heart and is thus the only zebrafish LGI homolog clearly expressed outside of neural tissues.

Without current knowledge of mutant phenotypes, the precise function of LGI genes in the embryo remains uncertain. It is interesting to note that LGI genes, particularly LGI1, are predominantly expressed in neural tissue. The LRRs present in LGI proteins have highest similarity to those found in the Slit protein family, which is involved in growth cone and neuronal guidance, and in Trk, a protein family thought to bind nerve growth factors and neurotrophins (reviewed in Kalachikov et al. 2002). Based upon the strong expression of lgi1b in cells underneath, and possibly derived from the rhombic lip, lgi1b is likely to play a role in neuronal cells migrating out of the proliferative zone in the lower rhombic lip toward their final location in the ventroanterior hindbrain (Koster and Fraser 2001).

Enhanced Purifying Selection in the LGI1 Gene Family

Wilson, Carlson, and White (1977) pioneered the idea that proteins with essential functions evolve more slowly, possibly due to stronger purifying selection. By comparing two genomes, several studies have indeed found either weak (Yang, Gu, and Li 2003) or strong (Hirsh and Fraser 2001, 2003; Jordan et al. 2002; Castillo-Davis and Hartl 2003; Wall et al. 2005; Zhang and He 2005) correlation between essential (disease-related) genes and rate of evolution. However, purifying selection is not unequivocally accepted by some as the reason for this correlation (Hurst and Smith 1999), and a few studies have identified other parameters that play either additional or more important roles in protein evolution, including overall gene expression rate and number of paralogs (Pal, Papp, and Hurst 2003; Yang, Gu, and Li 2003; Rocha and Danchin 2004). A recent paper that uses more sophisticated analytical methods concludes that “the correlation between gene dispensability and evolutionary rate, although low, is highly significant” (Zhang and He 2005). In particular, Thomas et al. (2003) have shown that cancer-related genes experience significantly stronger purifying selection than other disease genes and nondisease genes, as indicated by $K_{s}$/$K_{a}$ values over the entire sequence of orthologous proteins. However, it is possible that such a comparably refined method to calculate evolutionary pressure results in
an underestimate of disease genes under purifying selection. More sophisticated models, in which functional subdomains of proteins or even single amino acids are scanned rather than the entire protein, may reveal purifying selection that may be masked by a majority of neutral mutations in less important domains.

We therefore tested vertebrate LGI genes from mammals and teleosts for signs of natural positive or negative selection in coding regions at the level of individual amino acids. Interestingly, LGI1 and LGI4 orthologs show evidence for strong negative natural selection (purifying selection), while the remaining groups of LGI orthologs exhibited rather moderate signs of negative selection pressure (see \( \omega \) values, fig. 1C). Purifying selection is the form of natural selection that acts to eliminate selectively deleterious replacement mutations. In this sense, it might counteract mutations that have deleterious effects on protein function. Using the PAML software (Yang 1997) we assigned three classes of selection pressure within the LGI proteins, including two classes of highly conserved and constrained residues and one class of more relaxed residues (Fig. S1, Supplementary Material online). By performing a chi-square test, using the Statistica software, we found that ADLTE mutations predominantly occurred in the most constrained sites rather than being randomly dispersed within the protein (\( \chi^2 = 6.083; \ P < 0.05 \)).

Expression and selection data demonstrate that LGI1 and its orthologs differ from LGI2 and LGI3. Unfortunately, a clear statement for LGI4 is not possible, as the gene is absent in fish and no embryonic expression data are available to date in any other model organism. We have shown that gene expression between paralogous zebrafish lgi genes differs quite remarkably, which is in agreement with observations from a large number of duplicated genes (e.g., Huminiecki and Wolfe 2004; Rastogi and Liberles 2005). In contrast, when truly orthologous genes are compared between species, their expression patterns can show a considerable degree of conservation.

The expression of LGI genes in mammalian embryos has not yet been examined. To address the point if lgi1 expression patterns are conserved between zebrafish and mouse, we have compared lgi1 expression between the adult zebrafish and mouse brains (Kalachikov et al. 2002) and at this level of resolution do find clear similarities in lgi1 expression between both species. The high expression of the lgi1 genes in zebrafish CNS and high levels of purifying selection among the LGI1 genes in vertebrates argue for an essential role of this gene in developmental or physiological processes of the brain. Our data therefore show that mutations in LGI1 have a high a priori probability to be pathogenetic, a prediction which has already proven to be true. The neuronal expression of the remaining LGI genes is mostly restricted to a few cells, and, although under purifying selection, they are less constrained than LGI1 genes. However, because the expression patterns for mammalian LGI2 and LGI3 are not known and strong purifying selection was not detected for these genes, our results are of only limited value to predict or reject an involvement of these genes in diseases.

More generally, we propose that the approach outlined in this paper will be useful in selecting those genes from a larger gene family for further functional characterization that can be expected to be indispensable. In an initial simple procedure, orthologous genes from different organismal groups would be identified and assayed for evolutionary pressures using the PAML software. An estimated \( \omega \) ratio close to zero will be indicative of essential genes so that subsequent expression analyses can be targeted toward such disease candidates.

**Supplementary Material**

The Supplementary Data File which contains Supplementary Materials and Methods section, Supplementary Figure S1, Supplementary Tables S1–S5, and Supplementary References is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We thank I. May and K. Hoffmann for technical assistance and J. Freudenberg, S. Mercurio, and M. Mione for helpful discussions. This work was supported by a grant from the Nationales Genomforschungszentrum 2 (NGFN2) to O.K.S., a Landesgraduiertenstipendium to Y.G., grants from the Deutsche Forschungsgemeinschaft to A.M., and funding from Konstanz University to A.M. and G.B.

**Literature Cited**


.. [2216] Gu et al.


Douglas Crawford, Associate Editor

Accepted July 7, 2005