

Published in final edited form as:

Gene. 2004 October 27; 341: 255–266. doi:10.1016/j.gene.2004.07.003.

Representation of an immune responsive gene family encoding fibrinogen-related proteins in the freshwater mollusc *Biomphalaria glabrata*, an intermediate host for *Schistosoma mansoni*

Si-Ming Zhang and Eric S. Loker*

Department of Biology, University of New Mexico, Albuquerque, NM 87131, USA

Abstract

Fibrinogen-related proteins (FREPs) are found in the hemolymph of the freshwater snail *Biomphalaria glabrata*, are up-regulated following exposure to digenetic trematode parasites, and bind to trematode larval surfaces, suggestive of a role in internal defense. Southern blot and degenerate-polymerase chain reaction (PCR) analyses were undertaken to better understand the diversity of the FREP-encoding gene family. Probes corresponding to the N-terminal IgSF domains of specific *FREP* gene subfamilies (*FREPs* 2, 3, 4, 7, 12 and 13) revealed between 1 to 8 loci per subfamily on Southern blots. Probes representing the relatively conserved C-terminal fibrinogen domain of *FREPs* bound many sequences in Southern blots of genomic DNA from *B. glabrata*, and from two related gastropod species, *Biomphalaria pfeifferi* and *Helisoma trivolvis*. Using degenerate-PCR, we obtained 42 unique fibrinogen-encoding sequences from 180 clones derived from a single individual of the M-line strain of *B. glabrata*, further supporting the notion of their abundant representation in the *B. glabrata* genome. The fibrinogen-encoding sequences of *FREPs* encoding one or two IgSF domains tended to separate into distinct clades, but bootstrap support for this separation was low. A novel category of fibrinogen-encoding sequence was also revealed. This study provides the approximate number of gene copies in several *FREP* subfamilies, confirms the existence of a diverse *FREP* gene family, reports additional unusual sequences encoding fibrinogen-like molecules, and provides further justification to explore the functional roles of FREPs in both *B. glabrata* and *B. pfeifferi*, both important intermediate hosts of the human pathogen, *Schistosoma mansoni*.

Keywords

Innate immunity; Southern blot; Freshwater snail; Invertebrate; *Echinostoma paraensei*

1. Introduction

Fibrinogen-related proteins (FREPs) are present in the hemolymph of the freshwater gastropod *Biomphalaria glabrata*. They are known to be produced in hemocytes, the circulating defense cells of snails, and at least some categories of FREPs are up-regulated following infection with parasites like the digenetic trematode *Echinostoma paraensei*. Some FREPs are capable of binding to parasite surfaces and can precipitate soluble parasite antigens from solution, prompting speculation that they play a role in internal defense (Adema et al., 1997). The defense responses of *B. glabrata* are a relevant concern as this

snail is one of the most important intermediate hosts for another digenetic trematode, *Schistosoma mansoni*, a parasite that still infects about 83 million people (Crompton, 1999). The knowledge gained through the study of molecules or genes involved in snail response to trematode infection will help us to better understand the underlying mechanisms of the snail host and parasite interaction.

FREPs possess a unique molecular structure, having one or two immunoglobulin superfamily (IgSF) domains at the N terminus (referred to below as one-IgSF or two-IgSF FREPs) and a fibrinogen (FBG) domain at the C terminus (Léonard et al., 2001; Zhang et al., 2001). IgSF domains are best known from the vertebrate immune system but are also found in the nervous system of vertebrates and invertebrates (Du Pasquier and Flajnik, 1999; Du Pasquier, 2000). Several IgSF members have been described from invertebrates (Mendoza and Faye, 1999), including some functioning in the context of internal defense, such as hemolin in *Drosophila melanogaster* (Sun et al., 1990), and molluscan defense molecule (MDM) in *Lymnaea stagnalis* (Hoek et al., 1996).

Polypeptides containing fibrinogen domains have well-characterized functions in blood clotting and in innate immune responses in vertebrates. Our findings regarding FREPs from *B. glabrata* (Adema et al., 1997) led to the hypothesis that an ancestral function of fibrinogen molecules was in innate immunity. Shortly after the initial description of FREPs, several studies revealed fibrinogen-containing molecules functioning as innate-type defense factors in vertebrates and invertebrates (Kurachi et al., 1998; Gokudan et al., 1999; De Gregorio et al., 2001; Kenjo et al., 2001; Dehal et al., 2002; Zdobnov et al., 2002; Holmskov et al., 2003). Based on data currently available, fibrinogen-related molecules found in invertebrates function in innate immunity (Adema et al., 1997; Kurachi et al., 1998; Gokudan et al., 1999; Dimopoulos et al., 2000) and development (Baker et al., 1990; Xu and Doolittle, 1990).

We have subsequently learned that FREPs are encoded by a gene family, and we have uncovered several mechanisms for generating *FREP* diversity at the genomic DNA and mRNA levels. Thirteen subfamilies have been proposed in the *FREP* gene family (Léonard et al., 2001; Zhang et al., 2001; Zhang and Loker, 2003). In addition, it was suggested that *FREP* retrosequences and alternatively spliced *FREP* transcripts occur in *B. glabrata* (Zhang and Loker, 2003). Although additional work is required to fully unveil the functional role of FREPs, in aggregate our studies suggest that the *FREP* gene family can serve as a model for studying gene families involved in immune responses in invertebrates, mechanisms of diversification of immune related genes, and the molecular aspects of interactions between host and parasite.

Our previous *FREP* gene data were obtained largely using PCR techniques. As the complex diversity was revealed using PCR, it was necessary to relate this to more classical methods such as Southern hybridization to determine the extent of *FREP* gene diversity. For instance, we wanted to know how many loci are present in a particular *FREP* subfamily. Are the same numbers of *FREP* loci present in different strains of *B. glabrata*, or even in closely related species? Are there more fibrinogen-encoding sequences in the *B. glabrata* genome that are still unknown to us? What are the evolutionary relationships among the fibrinogen sequences?

To address the above questions, we first conducted Southern blot analyses on *B. glabrata* and related species using *FREP* subfamily-specific probes and probes conserved across *FREP* subfamilies. Next, we directly amplified fibrinogen-encoding regions of *FREPs* using degenerate-PCR. The data obtained provide insight into the nature of *FREP* diversity and greatly expand our understanding of the *FREP*-encoding gene family.

2. Materials and methods

2.1. Biological samples

Biomphalaria glabrata M-line, BS-90 and 13-16-R1 strains were maintained in the laboratory. *B. glabrata*, *Biomphalaria pfeifferi* and *Helisoma trivolvis* isolates were collected from Brazil (Barreiro), Kenya and USA (Albuquerque), respectively. Live specimens were used for extraction of DNA and RNA.

The *B. glabrata* embryonic (Bge) cell line was originally obtained from American Type Culture Collection (ATCC CRL 1494). The Bge cell line was established from embryonic cells that were in turn derived from multiple embryos at the trochophore to early shell stage of development (Hansen, 1976). Bge cells were maintained at 26 °C in complete Bge cell medium (Hansen, 1976), supplemented by 50 µg/µl gentamicin sulfate (Sigma, St. Louis, USA) and 5% fetal bovine serum (Sigma).

2.2. Extraction of genomic DNA and mRNA

Individual snails (not pooled samples) were the source of genomic DNA used in the present study for the purpose of Southern blot and PCR. Genomic DNA extraction from snails and Bge cells was carried out using the CTAB method (Winnepenninckx et al., 1993). The whole bodies of 30 juvenile snails (6–8 mm in shell diameter), either unexposed snails or snails exposed for 1 day to the digenetic trematode *E. paraensei*, were used for extraction of RNA (RNA isolation kit, Stratagene, La Jolla, CA, USA). mRNA was purified from total RNA using an oligo(dT) cellulose column method (Poly(A) Quik mRNA isolation kit, Stratagene). The concentration and purity of genomic DNA and mRNA were determined spectrophotometrically. The method for infection of snails with *E. paraensei* was described by Loker and Hertel (1987).

2.3. Generation of oligonucleotide probes

The choice of *FREP* subfamily-specific- or *FREP*-conserved probes was made based on available *FREP* sequence information; the IgSF regions are highly variable whereas the FBG region is relatively conserved (Léonard et al., 2001; Zhang et al., 2001; Zhang and Loker, 2003). For a particular *FREP* subfamily, we used IgSF region sequence as a subfamily-specific probe to examine the Southern pattern. Regarding the conserved probe, in order to hybridize as many FBG sequences as possible, we mixed FBGs from *FREP2* and *FREP3*, representing the two known types of *FREPs* (one-IgSF- and two-IgSF-*FREPs*, respectively). All oligonucleotide probes were amplified by PCR from complementary DNA (cDNA) except for the *FREP7*-specific probe generated from the exon 2 region of *FREP7* from genomic DNA. A schematic diagram of the general structure of the two known types of *FREPs*, the approximate locations of probes and abbreviations applied throughout the paper, are presented in Fig. 1. The size and location of the primers used for the generation of the oligonucleotide probes are presented in Table 1, and primer sequences are listed in Table 2.

2.4. Southern blot analyses

Southern blot analysis was conducted using a modified version of the SouthernMax™ method (Ambion, TX, USA). For each 500 µl restriction digestion, genomic DNA (~35 µg) was digested by the restriction enzymes (280–350 IU) at 37 °C for overnight (~17 h). Phenol and chloroform (1:1) were used to purify the digested genomic DNA. DNA was precipitated with ethanol. Digested DNA was dissolved in 20 µl distilled water. About 15–20 µg digested DNA was loaded on an 0.85% agarose gel (13 cm in length) and run at 22 V overnight (~17 h). DNA was transferred to a positively charged nylon membrane (BrightStar™-Plus, Ambion).

All probes used were radiolabelled with [α - 32 P]dATP using random priming (Strip-EZTM, Ambion). Pre-hybridization (42 °C, 30 min) and hybridization (42 °C, overnight) were performed using the ULTRAhybTM hybridization buffer which contains 50% formamide (Ambion). The blot was washed at 42 °C in low stringency buffer twice, then in high stringency buffer twice. The low stringency buffer and high stringency buffer are equivalent to 2 \times SSC, 0.1% SDS and 0.1 \times SSC, 0.1% SDS, respectively. Washing time was 10–20 min, depending on the quality of the blot and the number of times the blot was probed. For re-probing, the blot was washed twice with the probe degradation buffer provided by the supplier to remove previous hybridization signals (Strip-EZTM, Ambion).

2.5. Northern blot analysis

A total of 1.66 μ g mRNA from both exposed and unexposed snails were separated on a 1.0% agarose gel. Electrophoresis and RNA transfer were performed according to the supplier's instruction (NorthernMaxTM, Ambion). Hybridization, blot-washing and re-probing were performed in the same fashion as the Southern blot analysis described above. In addition to loading the same amount of mRNA, the staining intensity of 16S mitochondrial RNA (16SmtRNA) was used as a standard to assure that equal amounts of mRNA were loaded (Davidson and Swalla, 2002). Snail 16SmtRNA is A/T-rich and co-purified with poly-A mRNA when using the oligo(dT) cellulose column (Nowak et al., 2004).

2.6. Degenerate-PCR

In order to amplify potential fibrinogen sequences, we examined all fibrinogen sequences known from *FREPs* (Léonard et al., 2001; Zhang et al., 2001; Zhang and Loker, 2003) and designed two degenerate primers from the most conserved region of the *FREP* genes, a sense primer FbgFa and an anti-sense primer FbgRa. In addition, a set of non-degenerate sense primers FbgFb located in the FbgFa region was designed based on the majority of sequences used for primer design. Three primers (FbgFa, FbgFb and FbgRa) were added to the PCR reaction. The sequences of the three primers are presented in Table 2. The expected DNA fragments (excluding primer regions) amplified by this protocol would contain 18 canonical residues, 75% of all the canonical residues (24 in total) in the classic fibrinogen domains proposed by Doolittle (1992). The proofreading DNA polymerase, Platinum *Pfx* (Invitrogen, Carlsbad, CA), was used. PCR ingredients were added according to the supplier's instruction (Invitrogen). The template was genomic DNA from a single M-line *B. glabrata*. The PCR profile was as follows: 94 °C for 2 min for initial denaturation, 94 °C for 15 s, 50 °C for 30 s, 68 °C for 1 min 30 s for 33 cycles, then 68 °C for 5 min for final extension.

2.7. Cloning and sequencing

Amplicons were cloned into the pCR-Blunt II-TOPO vector (Invitrogen). The profile for sequencing reactions (BigDye, Applied Biosystems, Foster, CA) was: 96 °C for 17 s, 50 °C for 12 s, 60 °C for 4 min, with a total of 29 cycles. The extension products resulting from cycle sequencing were analyzed on an ABI 377 sequencer (Applied Biosystems).

2.8. Data analyses

Sequences were checked for accuracy using Chromas v 1.42 (<http://www.technelysium.com.au/chromas>). Alignments were performed in ClustalX 1.81 (Thompson et al., 1997) and visually checked using GeneDoc 2.6.002 (Nicholas and Nicholas, 1997). The percentage identities of nucleotide (nt) and amino acid (aa) sequences were calculated in BioEdit 5.0.9 (Hall, 1999). BLAST searches were performed using the NCBI BLAST program (<http://www.ncbi.nlm.nih.gov/blast>).

The phylogenetic relationships of fibrinogen sequences were determined using minimum evolution (ME) and maximum parsimony (MP) under P-distance in *PAUP** 4.0b10 (Swofford, 1998) and bootstrap support values after 1000 replicates were obtained for the ME tree. A fibrinogen-like sequence from *Limax flavus* sialic-acid-binding lectin 2 (SABL2, GenBank accession number: AF060450; Kurachi et al., 1998) was chosen as the outgroup.

3. Results

3.1. Enumeration of loci of six FREP subfamilies, estimated by subfamily-specific probes

Of 13 *FREP* subfamilies documented, six subfamilies, *FREPs 2, 3, 4, 7, 12* and *13*, which have been more fully characterized (Léonard et al., 2001; Zhang et al., 2001; Zhang and Loker, 2003), were investigated in the present study. The IgSF regions of the remaining *FREP* gene subfamilies (*FREPs 1, 5, 6, 8, 9, 10* and *11*) are poorly known, preventing us from undertaking Southern blot analysis. Genomic DNA from a single M-line snail was digested with four restriction enzymes, *EcoRI*, *HindIII*, *PstI* and *HaeIII*, and the ensuing blot was re-probed using *FREP*-specific probes generated from the IgSF region of the various *FREPs* (Fig. 2). Since we did not know whether the individual M-line snail used for Southern blot was homozygous at the various *FREP* loci, one locus may encode one or two sequences. Table 3 shows the approximate number of loci estimated for *FREPs 2, 3, 4, 7, 12* and *13* based on the strongly hybridizing bands. The number of genomic copies of *FREPs* that encode two tandem IgSF domains (*FREPs 3, 7, 12* and *13*) was in most cases higher than the number of copies for *FREPs* encoding a single IgSF domain (*FREPs 2* and *4*).

3.2. Variation in FREP subfamilies uncovered from different *B. glabrata* strains and related species

Additionally, snails of the 13-16-R1 and BS90 lab-adapted strains of *B. glabrata* and *B. glabrata* recently collected from the field in Brazil were examined. Fresh cells of the Bge cell line, another species of *Biomphalaria* (*B. pfeifferi*), and snails representing a second genus in the family Planorbidae, *H. trivolvis*, were included in Southern blot analyses (Fig. 3). For all strains of *B. glabrata*, the approximate numbers of loci for each *FREP* subfamily were generally consistent with that observed in the M-line snail (Fig. 2). For *H. trivolvis*, out of six subfamily-specific probes used, only the *FREP3*-specific probe hybridized with genomic DNA and the pattern of bands (three bands) was found to be similar to those for the *B. glabrata* strains. Southern blot analysis of *B. pfeifferi* obtained from the field in Kenya showed that both one-IgSF- and two-IgSF *FREPs* (*FREPs 2* and *3*, respectively) exist in this species (lane 10 in Fig. 3B). Unexpectedly, the number of *FREP3* bands obtained from Bge cells was almost twice the number observed in the *B. glabrata* strains, a phenomenon not noted in other *FREP* subfamilies.

3.3. Southern blot analyses using mixed fibrinogen probes confirmed the extraordinary diversity of the FREP family

To gain further insight into the diversity of the *FREP* family, we probed blots with a mixture of two fibrinogen probes (*FREP2*- and *FREP3*-fibrinogen). At least 24 visible bands were found in lanes loaded with DNA from the different strains of *B. glabrata*, Bge cells, *H. trivolvis*, *B. pfeifferi* and (Fig. 4A,B). These results confirm our previous notion that the region encoding the fibrinogen domain is relatively conserved across *FREP* subfamilies and can cross-hybridize to many *FREPs*, and suggest that the *FREP* gene family is indeed diverse in *B. glabrata*. In contrast, *H. trivolvis* shows fewer numbers of bands consistent with data derived from *FREP*-specific probes, in which only the *FREP3* subfamily-specific probe shows visible bands (see Fig. 3).

3.4. Northern blot analysis shows different subfamilies may be expressed differentially

Expression profiles comparing unexposed snails and snails exposed for one day to the trematode *E. paraensei* were generated using *FREPs* 2, 3 and 12 subfamily-specific probes, and the mixed fibrinogen probes previously applied in Southern analyses. The objective of Northern analysis was to compare *FREP* expression levels pre- and post-exposure. Northern blot analysis using the *FREP2*-specific probe revealed the strongest up-regulation post-exposure to *E. paraensei*. *FREPs* 3 and 12 showed a more moderate degree of up-regulation (Fig. 5), consistent with data generated by quantitative-PCR (Hertel et al., unpublished data). Interestingly, in Northern blot analysis, the band that hybridized with the mixed fibrinogen probes (*FREP2*-fibrinogen included) was obviously less intense than the band which hybridized with the *FREP2*-specific probe (i.e., *FREP2*-IgSF probe). The lack of a band or the likely weak smear in the non-exposed sample probed by the conserved probe might have resulted from low expression of *FREPs*, repeated stripping of the blot, alternative splicing of *FREP* genes described previously (Zhang and Loker, 2003) or a combination thereof.

3.5. Forty-two fibrinogen sequences were obtained from a single M-line snail, including a novel category of fibrinogen sequence

Since Southern blot analyses unequivocally demonstrated that a large number of fibrinogen sequences were present in the *B. glabrata* genome, we undertook a degenerate-PCR experiment to survey the variety of fibrinogen sequences and to examine the relationships among the sequences. A total of 180 clones were obtained from a PCR amplification generated from genomic DNA from an individual M-line snail. From these clones, 42 unique fibrinogen sequences were obtained, 4 of which (BG119c, BG162t, BG11t and BG65t) were identical to the homologous fibrinogen regions of *FREPs* 2, 7, 3a and 3b, respectively (Léonard et al., 2001; Zhang et al., 2001; Zhang and Loker, 2003). The remaining 38 sequences (GenBank accession nos.: AY309257, AY559455-AY559491) were different from any fibrinogen sequences thus far reported. The identities of nucleotides and amino acids of the 42 fibrinogen sequences obtained were 47–99% and 47–100%, respectively (Table 4).

As shown by previous studies, *FREPs* 2 and 4 are one-IgSF *FREPs* (Léonard et al., 2001), whereas *FREPs* 3, 7, 12 and 13 are two-IgSF *FREPs* (Zhang et al., 2001; Zhang and Loker, 2003). Using the fibrinogen-encoding region as a basis of comparison, a phylogenetic tree was constructed using the method of ME. The tree shows that the two-IgSF *FREPs* (*FREPs* 3, 7, 12 and 13) fall into a clade separate from one-IgSF *FREPs* (*FREPs* 2 and 4) and other fibrinogen-encoding genes from snails, although bootstrap support for this clade was low (Fig. 6). The MP tree showed a similar pattern with the exception of the position of two sequences (BG80t and BG13t). This discrepancy may be explained by the fact that these two sequences are so distantly related to the other *FREP* sequences that their positions in the gene trees are prone to move (MP tree not shown; see below for detailed description of the two sequences). In addition, it can be seen that the total number of two-IgSF *FREPs* (31 different sequences) revealed was at least four times higher than that of one-IgSF *FREPs* (7 different sequences) (Fig. 6). This suggests that the diversity of two-IgSF *FREPs* exceeds that of one-IgSF *FREPs*, consistent with the Southern blot analyses described above (Figs. 2 and 3).

The distinctiveness of the two sequences described above (BG80t and BG13t) is further supported by the fact that they have 13 of the 18 canonical fibrinogen residues (Doolittle, 1992) predicted from the partial fibrinogen domains amplified here (Fig. 7). BLAST searches using deduced amino acid sequences provided support to indicate that BG13t and BG80t were indeed fibrinogen-encoding sequences. These two new sequences were 100% identical at the amino acid level, but clearly were less similar to the fibrinogen-encoding

sequences associated with *FREPs* (Table 4). To further explore this difference, we created a probe from one of these new sequences (BG13t) for hybridization with genomic DNA of Bge cells. Southern blots generated from Bge cells revealed two to three bands (Fig. 4C) that clearly differed from those revealed by known *FREP*-fibrinogen probes (Fig. 4A,B).

Taken together, our data suggest that BG13t and BG80t represent a novel category of fibrinogen sequence not previously documented in the *B. glabrata* genome. Two additional unusual fibrinogen sequences (BG106m and BG108t), which possess 16 and 15 canonical residues, respectively, were found (Fig. 7), and group more closely to most of the fibrinogen-encoding sequences of *FREPs* than do BG13t and BG80t (Table 4).

To facilitate submission of new sequences to GenBank, we assigned the particular sequence to a particular *FREP* subfamily based on the cluster into which it fell on the tree. For example, sequence BG36t was assigned as a member of *FREP3* subfamily because it fell into a clade containing other *FREP3* sequences. The temporary name of “fibrinogen-related domain” (FReD) was provided for those sequences with uncertain position in the phylogenetic tree.

4. Discussion

The Southern analyses presented here provide the first systematic documentation of the number of *FREP*-reactive bands in the genome of *B. glabrata*. Previous work reported that *FREPs* belong to a large gene family, and the criterion of 86% sequence identity has been used to define membership in a particular *FREP* subfamily (Zhang and Loker, 2003). As the probes designed to be subfamily-specific based on sequence data may nonetheless cross-react with members of other closely related subfamilies on Southern blots, despite of application of a high stringency washing buffer, caution is required in interpreting the number of loci in each subfamily. Many factors may affect the outcome of Southern hybridization. For example, different formulations for hybridization and washing buffers may yield different intensity of bands. In addition, we have no information regarding the homozygosity of *FREP* loci. Therefore, the number of loci we provide for each *FREP* subfamily should be considered as approximations, rather than an exact number of loci.

The present study provides evidence for the presence of *FREP*-like sequences from another *Biomphalaria* species from a different continent, namely *B. pfeifferi*, a widely distributed intermediate host for *S. mansoni* from sub-Saharan Africa. The sister genus of *Biomphalaria* is likely to be *Helisoma* (Morgan et al., 2002), and a representative of that genus, *H. trivolvis* from North America, was also shown to possess *FREP*-like sequences. Based on Southern blot analysis, *H. trivolvis* appears to have either fewer fibrinogen copies than *B. glabrata*, or the corresponding sequences have diverged significantly, resulting in decreased hybridization with the *B. glabrata* probes. A *FREP2*-like sequence has been cloned from *H. trivolvis* (Adema, personal communication), suggesting the latter explanation may apply. Moreover, 65-kDa soluble poly-peptides believed to be encoded by one-IgSF *FREPs*, like *FREP2* have been shown to be produced by *H. trivolvis* and to precipitate trematode secretory products in a similar manner to one-IgSF *FREPs* from *B. glabrata* (Adema et al., 1999).

Although we have obtained a large number of fibrinogen-encoding sequences from different gastropod species using PCR and cloning, the *B. glabrata*-derived fibrinogen probes applied in the present study failed to detect fibrinogen sequences in most of these species (Zhang and Loker, unpublished data), again suggesting that the level of homology with *B. glabrata* *FREPs* may not reach the level required for detection by Southern blot analysis. This may also suggest that *FREPs* have undergone extensive diversification. Two-IgSF *FREPs* may be more readily detected in species divergent from *B. glabrata* simply because they exist at

more loci. Alternatively, they may be more conserved, allowing the genes to persist within the closely related species.

This study has also revealed novel fibrinogen sequences encoded by two to three loci based on Southern blot, phylogenetic and sequence analyses (Figs. 4, 6 and 7 and Table 4). They are unlike those lying C-terminal to the IgSF domains of FREPs. They may encode polypeptides that function as independent-fibrinogen domains that combine more than two IgSF domains or that are possibly associated with domains other than IgSF.

This study has employed two distinct methods to investigate fibrinogen diversity. First, based on our previous findings that the fibrinogen region is conserved, we used mixed fibrinogen probes to screen the genome of the *B. glabrata*. Southern blot analysis (Fig. 4) showed at least 24 visible bands, with other bands likely escaping our attention because of overlap. The *B. glabrata* genome is rich in fibrinogen sequences, a point reinforced by the use of the second method, a direct search for fibrinogen sequences from a single M-line snail using degenerate-PCR. A total of 42 unique fibrinogen sequences were found from 180 clones analyzed from this individual, but it is likely that even more fibrinogen sequences are present. This conclusion is based on the fact that only 180 clones were sequenced, degenerate primers may not be able to amplify all potential fibrinogen sequences (especially for those sequences where the primer sites have less homology to primer sequences), and that some alleles may have been lost due to the potential homozygosity of the M-line laboratory strain.

The Northern blot data presented here confirm an earlier observation of up-regulated *FREP* expression following exposure to digenean infection (Adema et al., 1997). The lack of an obvious band in lanes loaded with mRNA from unexposed snails and probed with FBG probes does not imply that *FREP* expression does not occur in the absence of parasitism (Fig. 5). We know *FREPs* are expressed at a certain level in unexposed snails. This may be largely due to repeatedly stripping of Northern blot. The difference noted in the intensity of the transcript band in the lane probed with a *FREP2*-IgSF probe and a mixed fibrinogen probe may be indicative of the alternative splicing that has been observed in *FREPs* (Zhang and Loker, 2003). As far as is known, all alternatively spliced *FREP* transcripts contain IgSF regions and some alternatively spliced forms lack FBG region. For a particular *FREP* subfamily, a subfamily-specific probe developed from IgSF sequences can therefore detect all transcripts including alternatively spliced forms, whereas FBG probes could hybridize only to those transcripts that contain intact FBG sequences. The relative abundance of alternatively spliced forms and the mechanism(s) by which they are generated are presently unknown. It will be of interest to learn if alternatively spliced transcripts are associated with particular stresses or challenges.

Bge cells have been widely used in molecular studies of *B. glabrata* biology (Davids et al., 1999; Duclermortier et al., 1999; Humphries et al., 2001), and it has been noted that Bge cells share several features with the circulating defense cells (hemocytes) involved in molluscan immune defense (Duclermortier et al., 1999). Our Southern blot analysis demonstrated that Bge cells, the only cell line available for *B. glabrata* or, for that matter, any other mollusc, contain twice the normal number of *FREP3* gene copies. The duplication may have occurred in either an entire chromosome carrying the *FREP3* gene or in a chromosomal region in which *FREP3* loci are located. Our results suggest that genetic alterations in the cells have occurred in the nearly 30-year period since they were first derived (Hansen, 1976). Additional work is needed to more fully characterize the genetic content of Bge cells.

Several intriguing questions arise from this analysis. What are the respective histories of the fibrinogen and IgSF domains, and when in their evolutionary histories did they first associate to form the single polypeptide that we have identified as a FREP? How did the *FREP* gene family further diversify? Also, for both fibrinogen and IgSF domains, for what other roles have they been co-opted in snails or other invertebrates? For example, 58 fibrinogen sequences were recently identified in the genome in *Anopheles gambiae* (Zdobnov et al., 2002), at least some of which function in immune responses (Dimopoulos et al., 2000; Christophides et al., 2002). Similarly, 13 fibrinogen sequences, some involved in immunity, occur in *Drosophila* (Adams et al., 2000; De Gregorio et al., 2001). Even though we have barely begun to sample the *B. glabrata* genome, *FREPs* already comprise the largest fibrinogen gene family documented so far. The *FREP* gene family from *B. glabrata* is thus an intriguing system to further address the functional roles of fibrinogen molecules in invertebrates. Our Northern blot data suggests that one-IgSF *FREPs* such as *FREP2* are more responsive to digenetic trematode exposure than two-IgSF *FREPs*. What then might be the function of two-IgSF *FREPs*? Do they respond to other kinds of parasitic or antigenic stimuli or do they have totally different functions? Much remains to be investigated with respect to the genomic organization and functions of the *FREP* gene family.

Acknowledgments

We would like to thank our colleagues Mr. T. Nowak, Ms. L. Hertel, Dr. R. DeJong and Dr. C. Adema for their valuable help during the laboratory work and manuscript preparation, and Dr. T. Yoshino, University of Wisconsin, for assistance with Bge cell culture. This work was supported by NIH Grant (AI24340) and NIH Grant Number RR-1P20RR18754 from the Institutional Development Award (IDeA) Program of the National Center for Research Resources. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NIH.

Abbreviations

aa	amino acid
nt	nucleotide
cDNA	complementary DNA
FBG	fibrinogen
FREP	fibrinogen-related protein
ICR	interceding region
IgSF	immunoglobulin superfamily
PCR	polymerase chain reaction
FReD	fibrinogen-related domain
ME	minimum evolution

References

- Adams MD, et al. The genome sequence of *Drosophila melanogaster*. *Science*. 2000; 287:2185–2195. [PubMed: 10731132]
- Adema CM, Hertel LA, Miller RD, Loker ES. A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc Natl Acad Sci U S A*. 1997; 94:8691–8996. [PubMed: 9238039]
- Adema CM, Hertel LA, Loker ES. Evidence from two planorbid snails of a complex and dedicated response to digenetic (echinostome) infection. *Parasitology*. 1999; 119:395–404. [PubMed: 10581618]

- Baker NE, Mlodzik M, Rubin GM. Spacing differentiation in the developing *Drosophila* eye: a fibrinogen-related lateral inhibitor encoded by *scabrous*. *Science*. 1990; 250:1370–1377. [PubMed: 2175046]
- Christophides GK, et al. Immunity-related genes and gene families in *Anopheles gambiae*. *Science*. 2002; 298:159–165. [PubMed: 12364793]
- Crompton DWT. How much human helminthiasis is there in the world? *J Parasitol*. 1999; 85:397–403. [PubMed: 10386428]
- Davids BJ, Wu XJ, Yoshino TP. Cloning of a β intergrin subunit cDNA from an embryonic cell line derived from the freshwater mollusc, *Biomphalaria glabrata*. *Gene*. 1999; 228:213–223. [PubMed: 10072774]
- Davidson B, Swalla BJ. A molecular analysis of ascidian metamorphosis reveals activation of an innate immune response. *Development*. 2002; 129:4739–4751. [PubMed: 12361966]
- De Gregorio E, Spellman PT, Rubin GM, Lemaitre B. Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc Natl Acad Sci U S A*. 2001; 98:12590–12595. [PubMed: 11606746]
- Dehal P, et al. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science*. 2002; 298:2157–2167. [PubMed: 12481130]
- Dimopoulos G, Casavant TL, Chang S, Scheetz T, Roberts C, Donohue M, Schultz J, Benes V, Bork P, Ansong W, Soares MB, Kafatos FC. *Anopheles gambiae* pilot gene discovery project: identification of mosquito innate immunity genes from expressed sequence tags generated from immune-competent cell lines. *Proc Natl Acad Sci U S A*. 2000; 97:6619–6624. [PubMed: 10841561]
- Doolittle RF. A detailed consideration of a principal domain of vertebrate fibrinogen and its relatives. *Protein Sci*. 1992; 1:1563–1577. [PubMed: 1304888]
- Duclermortier P, Lardans V, Serra E, Trottein F, Dissous C. *Biomphalaria glabrata* embryonic cells express a protein with a domain homologous to the lectin domain of mammalian selectins. *Parasitol Res*. 1999; 85:481–486. [PubMed: 10344542]
- Du Pasquier, L.; Flajnik, M. Origin and evolution of the vertebrate immune system. In: Paul, WE., editor. *Fundamental Immunology*. 4. Lippincott-Raven Publishers; Philadelphia: 1999. p. 605–649.
- Du Pasquier, L. The phylogenetic origin of antigen-specific receptors. In: Du Pasquier, L.; Litman, GW., editors. *Origin and Evolution of the Vertebrate Immune System*. Springer-Verlag; Berlin: 2000. p. 159–189.
- Gokudan S, Muta T, Tsuda R, Koori K, Kawahara T, Seki N, Mizunoe Y, Wai SN, Iwanaga S, Kawabata SI. Horseshoe crab acetyl group-recognizing lectins involved in innate immunity are structurally related to fibrinogen. *Proc Natl Acad Sci U S A*. 1999; 96:10086–10091. [PubMed: 10468566]
- Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*. 1999; 41:95–98.
- Hansen, EL. A cell line from embryos of *Biomphalaria glabrata* (Plumonata): establishment and characteristics. In: Maramorosch, K., editor. *Invertebrate Tissue Culture: Research Applications*. Academic Press; New York: 1976. p. 75–97.
- Holmskov U, Thiel S, Jensenius J. Collectins and ficolins: humoral lectins of the innate immunity defense. *Annu Rev Immunol*. 2003; 21:547–578. [PubMed: 12524383]
- Hoek RM, Smit AB, Frings H, Vink JM, deJong-Brink M, Geraerts WPM. A new Ig-superfamily member, molluscan defence molecule (MDM) from *Lymnaea stagnalis*, is down-regulated during parasitosis. *Eur J Immunol*. 1996; 26:939–944. [PubMed: 8625992]
- Humphries JE, Elizondo L, Yoshino TP. Protein kinase C regulation of cell spreading in the molluscan *Biomphalaria glabrata* embryonic (Bge) cell line. *Biochim Biophys Acta*. 2001; 1540:243–252. [PubMed: 11583819]
- Kenjo A, Takahashi M, Matsushita M, Endo Y, Nakata M, Mizuochi T, Fujita T. Cloning and characterization of novel ficolins from the solitary ascidian, *Halocynthia roretzi*. *J Biol Chem*. 2001; 276:19959–19965. [PubMed: 11259430]

- Kurachi S, Song Z, Takagaki M, Yang Q, Winter HC, Kurachi K, Goldstein IJ. Sialic-acid-binding lectin from the slug *Limax flavus*: cloning, expression of the polypeptide, and tissue location. *Eur J Biochem.* 1998; 254:217–222. [PubMed: 9660173]
- Léonard PM, Adema CM, Zhang SM, Loker ES. Structure of two *FREP* genes that combine IgSF and fibrinogen domains, with comments on diversity of the *FREP* gene family in the snail *Biomphalaria glabrata*. *Gene.* 2001; 269:155–165. [PubMed: 11376947]
- Loker ES, Hertel LA. Alteration in *Biomphalaria glabrata* plasma induced by infection with the digenetic trematode *Enchiostoma paraensei*. *J Parasitol.* 1987; 75:503–513. [PubMed: 3598800]
- Mendoza HL, Faye I. Physiological aspects of the immunoglobulin super-family in invertebrates. *Dev Comp Immunol.* 1999; 23:359–374. [PubMed: 10426428]
- Morgan JAT, DeJong RJ, Jung YH, Khallaayoune K, Kock S, Mkoji GM, Loker ES. A phylogeny of planorbid snails, with implications for the evolution of *Schistosoma* parasites. *Mol Phylogenet Evol.* 2002; 25:477–488. [PubMed: 12450752]
- Nicholas, KB.; Nicholas, HB, Jr. GeneDoc: A Tool For Editing And Annotating Multiple Sequence Alignments. 1997. Distributed by the authors
- Nowak TS, Woodards AC, Jung Y, Adema CM, Loker ES. Identification of transcripts generated during the response of resistant *Biomphalaria glabrata* to *Schistosoma mansoni* infection using suppressive subtractive hybridization. *J Parasitol.* 2004 (in press).
- Sun SC, Lindstrom I, Boman HG, Faye I, Schmidt O. Hemolin: an insect-immune protein belonging to the immunoglobulin superfamily. *Science.* 1990; 250:1729–1732. [PubMed: 2270488]
- Swofford, DL. Phylogenetics Analysis Using Parsimony (* and Other Methods). Version 4.0b. Sinauer Associates; Sunderland, MA: 1998. *PAUP**.
- Thompson JD, Higgins DG, Gibson TJ. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 1997; 25:4876–4882. [PubMed: 9396791]
- Winnepenninckx B, Backeljau T, DeWachter R. Extraction of high-molecular-weight DNA from mollusks. *Trends Genet.* 1993; 9:407. [PubMed: 8122306]
- Xu X, Doolittle RF. Presence of a vertebrate fibrinogen-like sequence in an echinoderm. *Proc Natl Acad Sci U S A.* 1990; 87:2097–2101. [PubMed: 2315305]
- Zdobnov EM, et al. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science.* 2002; 298:149–159. [PubMed: 12364792]
- Zhang SM, Léonard PM, Adema CM, Loker ES. Parasite-responsive IgSF members in the snail *Biomphalaria glabrata*: characterization of novel genes with tandemly arranged IgSF domains and a fibrinogen domain. *Immunogenetics.* 2001; 53:684–694. [PubMed: 11797103]
- Zhang SM, Loker ES. The *FREP* gene family in the snail *Biomphalaria glabrata*: additional members, and evidence consistent with alternative splicing and *FREP* retrosequences. *Dev Comp Immunol.* 2003; 27:175–187. [PubMed: 12590969]

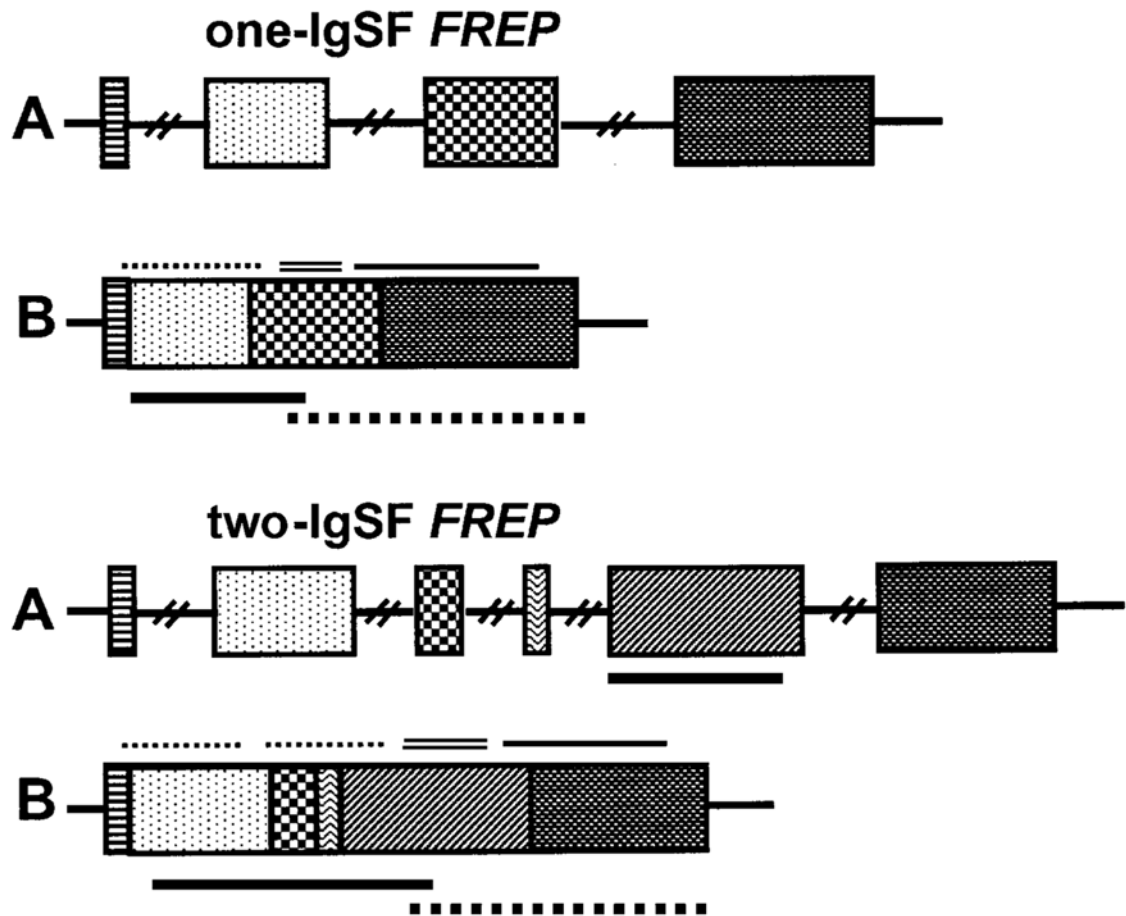


Fig. 1. Structures of two known types of *FREPs*, the one-IgSF *FREP* (upper) and the two-IgSF *FREP* (lower), are shown. (A) Genomic structure and (B) mRNA transcript. The solid line and dotted line below the mRNA transcript in (B) represent the approximate location of *FREP* subfamily-specific probes and *FREP* conserved probes, respectively. The exact length for each probe is described in Table 2. The dotted line(s), double line and solid line indicated above the mRNA transcripts represent the IgSF(s), ICR and FBG regions, respectively. Abbreviations: IgSF, Immunoglobulin superfamily; ICR, interceding region; FBG, fibrinogen. Textured boxes represent different exons.

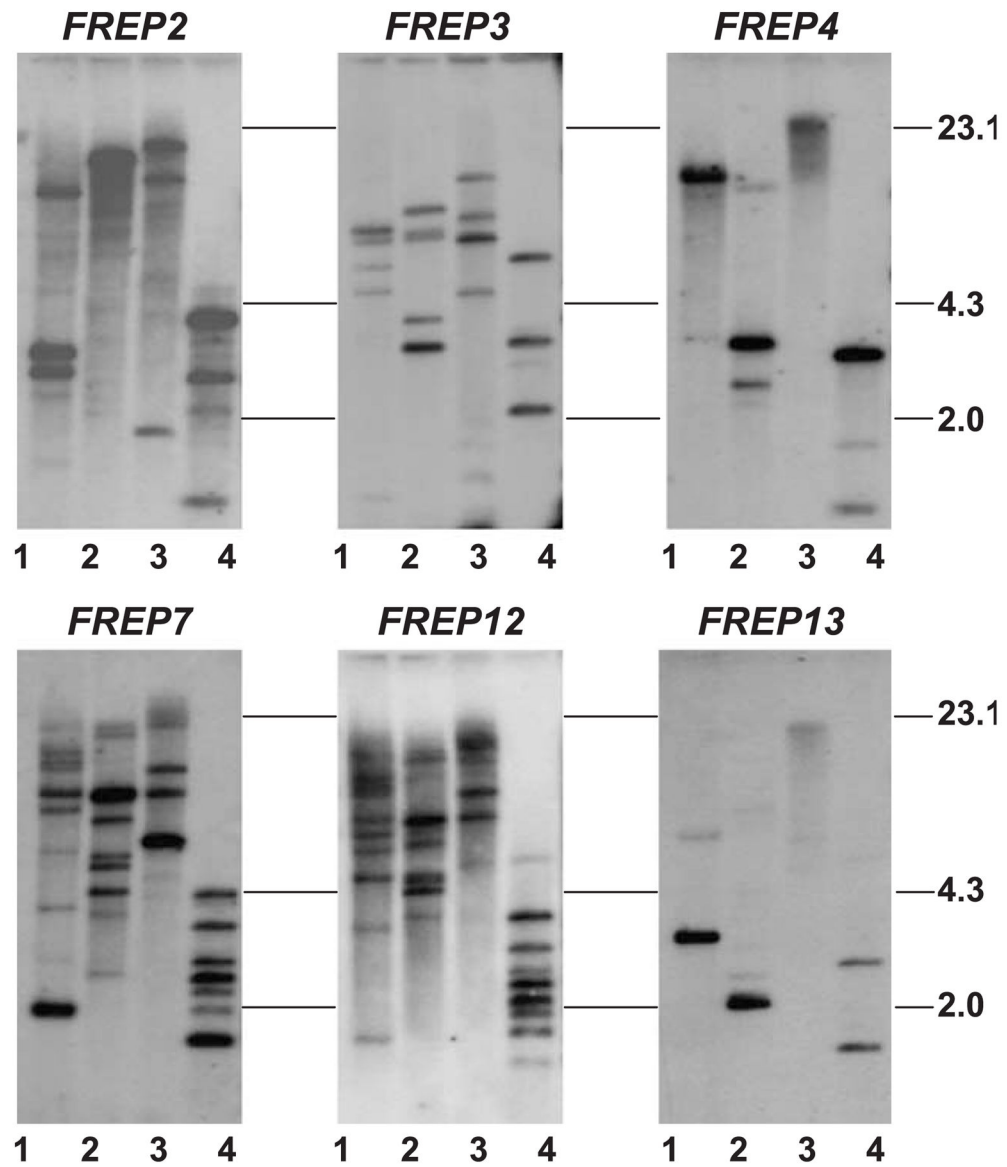


Fig. 2. Southern blots depicting the hybridization of *FREP* subfamily-specific probes to genomic DNA obtained from one individual of the M-line strain of *B. glabrata*. All six images are derived from the same membrane hybridized with different probes. The order for re-probing is *FREPs* 3, 2, 13, 4, 7 and 12. In each image, the patterns from lanes 1 to 4 show the genomic DNA digested with *EcoRI*, *HindIII*, *PstI* and *HaeIII*, respectively. Size markers in kilobases (kb) are indicated on the right.

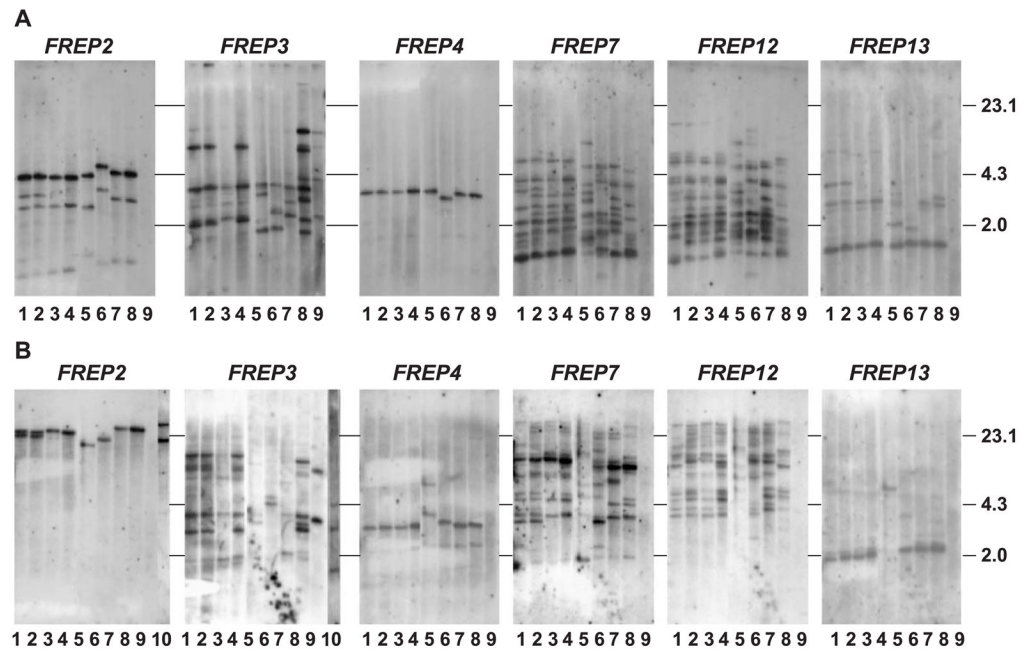


Fig. 3. Southern blots obtained from genomic DNA from different strains of *B. glabrata*, Bge cells, *B. pfeifferi* and *H. trivolvis*, digested with *Hae*III (A) or *Hind*III (B). In each image, lanes 1–4 are *B. glabrata* M-line and lanes 5–7 are *B. glabrata* Brazilian wild isolate, BS90, 13-16-R1, and lanes 8 and 9 are Bge cells and *H. trivolvis*, respectively. Lane 10 shows the genomic DNA of *B. pfeifferi* digested by *Hind*III and probed with *FREPs* 2 and 3 subfamily-specific probes. The order for re-probing for both restriction enzymes was *FREPs* 3, 2, 12, 4, 7 and 13.

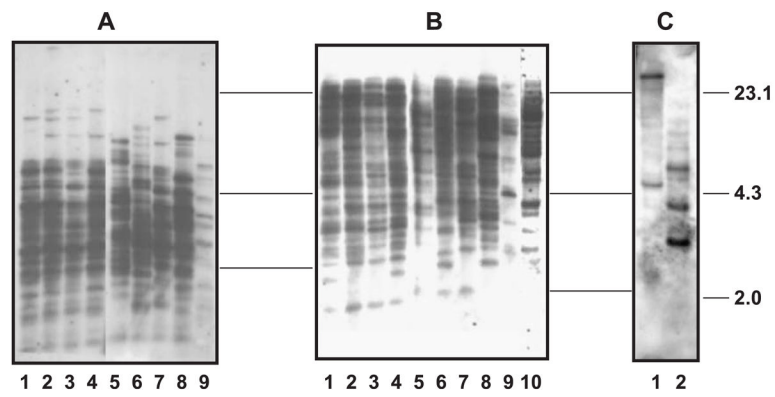


Fig. 4. Southern blot obtained from genomic DNA digested by *Hae*III (A) or *Hind*III (B) hybridized with mixed fibrinogen probes (*FREP2*- and *FREP3*-fibrinogen). In both images, the DNA samples from lanes 1 to 10 are same as those of Fig. 3. (C) Genomic DNA from Bge cells digested with *Hind*III (1) and *Hae*III (2) probed by the novel fibrinogen probe BG13t.

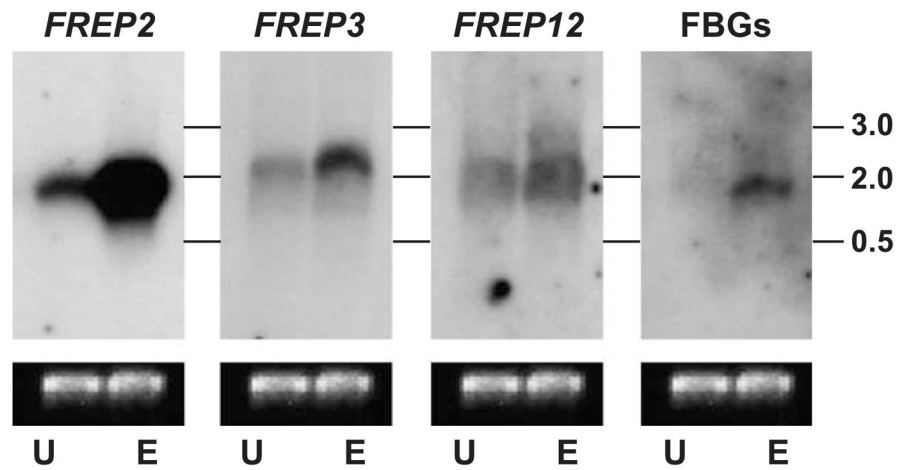


Fig. 5. Expression profile of M-line *B. glabrata* showing unexposed (U) versus exposed (1-day post-exposure to *E. paraensei*; E) using *FREP* subfamily-specific probes, or a mixed fibrinogen probe in Northern blot analysis. All images were produced from the same Northern blot hybridized with different probes. Size markers in kilobases (kb) are indicated on the right. The lower lane shows the staining intensity of 16SmtRNA from the same image. The order for re-probing was *FREPs* 2, 3, 12 and FBGs.

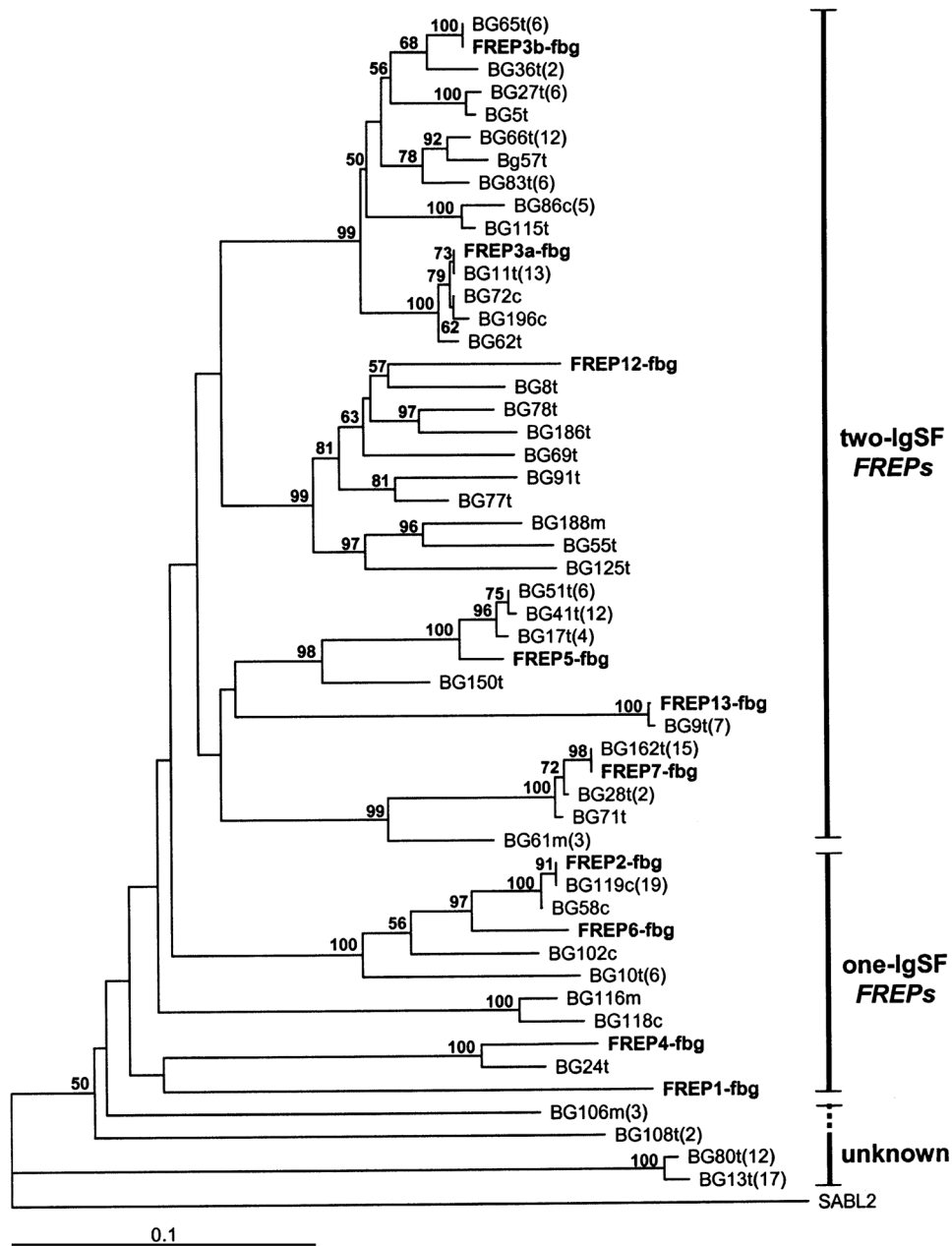


Fig. 6. Phylogenetic tree generated by ME showing the relationship of fibrinogen-encoding sequences. The bootstrap value at the each branch represents 1000 replicates using *PAUP4.0b* software. Bootstrap values less than 50% are not indicated. The tree is composed of 42 fibrinogen sequences obtained from genomic DNA from a single snail. *FREP* sequences we have previous reported are indicated in bold. A fibrinogen-like sequence of *L. flavus* sialic acid-binding lectin 2 (SABL2; GenBank accession number: AF060450) was used as the outgroup. The numbers in parentheses indicate the number of clones containing the sequence.

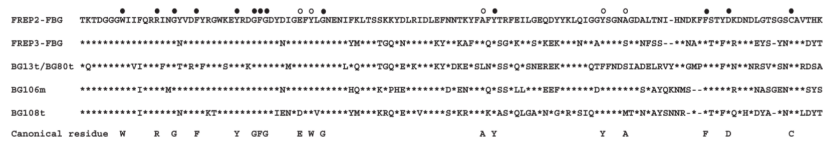


Fig. 7. Alignment of the corresponding fibrinogen (FBG) regions of *FREPs 2 and 3*, and the deduced amino acid sequences obtained from clones BG13t, BG80t, BG106m and BG108t. BG13t and BG80t encode for an identical amino acid sequence. Asterisks represent identical residues and dashes indicate gaps. The original 18 canonical fibrinogen residues proposed by Doolittle (1992) are listed in the last row. The canonical residues are indicated at the top by filled circles (present in all five sequences) or open circles (replacement in one or more sequences).

Table 1

General features of the oligonucleotide probes

Probe	Primers used for the generation of the probe	An approximate region the probe spans	Size of probe (nt)	Type of template
<i>FREP2</i> -specific	F2GSP17, F2PW1	IgSF and ICR	554	<i>FREP2</i> cDNA
<i>FREP4</i> -specific	F4GSP11b, F4GSP12b	IgSF and ICR	629	<i>FREP4</i> cDNA
<i>FREP3</i> -specific	F3GSP23a, F3GSP36a	IgSF1, IgSF2 and ICR	1253	<i>FREP3</i> cDNA
<i>FREP7</i> -specific	F7GSP13a, F7GSP26	IgSF1 and a small part of IgSF2	548	genomic DNA
<i>FREP12</i> -specific	F12PW4, F12PW5	IgSF1 and IgSF2	839	<i>FREP12</i> cDNA
<i>FREP13</i> -specific	F13GSP4b, F13GSP12	IgSF1 and a partial IgSF2	480	<i>FREP13</i> cDNA
FBG1	F2PW2, F2GSP18	FBG of <i>FREP2</i>	561	<i>FREP2</i> cDNA
FBG2	F3GSP4a, F3GSP27	IgSF2, ICR and FBG of <i>FREP3</i>	1261	<i>FREP3</i> cDNA
FBG3	FbgFb, FbgRa	Novel FBG sequence	436	genomic DNA

Abbreviations that appear in this table are given in Fig. 1.

Table 2

Primers used for the generation of the oligonucleotide probes and for degenerate-PCR

Primer	Sequence (5'-3')
F2GSP17	AAAATGGCGTCGCTACCACTTCGACTTG
F2PW1	TCTGGTAGCGTGGGACTGGCTCT
F4GSP11b	TGCTTGTTTCTAGTATCTGCAACTCTAG
F4GSP12b	GACGTCGGGTTTCGCCAGACATTCGC
F3GSP23a	ATGTCCAACCAAACGTCATATCGCCAGAG
F3GSP36a	GGTCTGATTTTTCACTTTTATCGAGGCAG
F7GSP13a	TCAGAACTTGTCATTGATGTCACCCAG
F7GSP26	TACTTGTCTATCATTACTTGCTAAAGAAC
F12PW4	ACCAAGTTCAACATCTGGATG
F12PW5	GATCTCCATCAGGTCCCATG
F13GSP4b	CTCGTTATAATGAACTATCAGAGACTTTG
F13GSP12	CATGAAGAATGTACAAAGACTGAAGTGAG
F2PW2	CTACCAGAGTCTTGTCGTGACG
F2GSP18	CGTAAGCCTGGGTTTTCCCACTTAATCTAG
F3GSP4a	CTATGACAACGCAATTAAGATTCC
F3GSP27	CAATTCTTAAATCAAAAAGTCCAGCAGAAG
FbgFa	TCYGGKTYARARGTCATGTGTGAC
FbgFb	TCCGGGTAAAGGTCATGTGTGAC
FbgRa	CAWBWKYNGTACCACCAGGCTCC

K=G/T, Y=C/T, R=A/G, W=A/T, B=G/T/C, N=A/T/G/C.

Table 3

The number of loci estimated in six *FREP* subfamilies

Subfamily	<i>EcoRI</i>		<i>HindIII</i>		<i>PstI</i>		<i>HaeIII</i>		Number of loci
	Homo	Heter	Homo	Heter	Homo	Heter	Homo	Heter	
<i>FREP2</i>	3	2	1	1	3	2	3	2	1-3
<i>FREP3</i>	3	2	5	3	4	2	3	2	2-5
<i>FREP4</i>	1	1	2	1	1	1	3	2	1-3
<i>FREP7</i>	7	4	8	4	>4	>2	7	4	3-8
<i>FREP12</i>	8	4	7	4	5	6	7	4	4-8
<i>FREP13</i>	1	1	1	1	unknown	unknown	2	1	1-2

Number of loci was estimated considering both homozygosity (homo) and heterozygosity (heter). number of loci = number of bands/2. Each band represents one allele, assuming no overlap of DNA sequences in a band. In some lanes, the exact numbers of bands were finally determined after the consideration of additional Southern data, which are not shown in this paper.

Table 4

Pairwise comparison of nucleotide (nt) and amino acid (aa) identities between different types of homologous FBG sequences

Pairwise comparison	nt identity (%)	aa identity (%)
All FBGs (42 FBGs+known FBG of <i>FREPs</i>)	47–99	50–100
BG13t vs. BG80t	99	100
All FBG excluding BG13t and BG80t	57–99	57–100
BG13t and BG80t vs. All FBGs excluding BG13t and BG80t	47–61	50–61
BG106m vs. BG13t and BG80t	54	50
BG106m vs. All FBG excluding BG13t and BG80t	52–65	58–71
BB108t vs. BG13t and BG80t	50	50
BG108t vs. All FBG excluding BG13t and BG80t	53–65	57–65
SABL2 vs. BG13t and BG80t	50	43
SABL2 vs. All FBG excluding BG13t and BG80t	54–59	47–55
SABL2 vs. BG106m	59	55
SABL2 vs. BG108t	52	49

SABL2, a fibrinogen-like sequence of *L. flavus* sialic acid-binding lectin 2; FBG, fibrinogen. Calculations were performed on values obtained from the alignment described in Section 2.8.