Opsin Repertoire and Expression Patterns in Horseshoe Crabs: Evidence from the Genome of Limulus polyphemus (Arthropoda: Chelicerata)

Barbara-Anne Battelle1,*, Joseph F. Ryan2, Karen E. Kempler1, Spencer R. Saraf1,7, Catherine E. Marten1,8, Wesley C. Warren3, Patrick J. Minx3, Michael J. Montague3, Pamela J. Green4, Skye A. Schmidt4, Lucinda Fulton3, Nipam H. Patel5, Meredith E. Protas5,9, Richard K. Wilson3 and Megan L. Porter6

1Whitney Laboratory for Marine Bioscience, Department of Neuroscience and Biology, University of Florida
2Whitney Laboratory for Marine Bioscience, Department of Biology, University of Florida
3McDonnell Genome Institute, Washington University School of Medicine in St. Louis
4Department of Plant and Soil Sciences, School of Marine Science and Policy, Delaware Biotechnology Institute, University of Delaware
5Department of Molecular Cell Biology, Center for Integrative Genomics, University of California, Berkeley
6Department of Biology, University of Hawaii, Manoa
7Present address: School of Marine and Atmospheric Sciences, Stony Brook University, Stony Brook, NY
8Present address: Department of Molecular Genetics and Microbiology, College of Medicine, University of Florida, Gainesville, FL
9Present address: Department of Natural Sciences and Mathematics, Dominican University of California, San Rafael, CA

*Corresponding author: E-mail: battelle@whitney.ufl.edu.

Accepted: April 24, 2016

Data deposition: This project has been deposited at NCBI GenBank under the accession PRJNA20489. A convenient portal for BLASTing and downloading the genome data is: http://ryanlab.whitney.ufl.edu/genomes/Lpol/.

Abstract

Horseshoe crabs are xiphosuran chelicerates, the sister group to arachnids. As such, they are important for understanding the most recent common ancestor of Euchelicerata and the evolution and diversification of Arthropoda. Limulus polyphemus is the most investigated of the four extant species of horseshoe crabs, and the structure and function of its visual system have long been a major focus of studies critical for understanding the evolution of visual systems in arthropods. Likewise, studies of genes encoding Limulus opsins, the protein component of the visual pigments, are critical for understanding opsin evolution and diversification among chelicerates, where knowledge of opsins is limited, and more broadly among arthropods. In the present study, we sequenced and assembled a high quality nuclear genomic sequence of L. polyphemus and used these data to annotate the full repertoire of Limulus opsins. We conducted a detailed phylogenetic analysis of Limulus opsins, including using gene structure and synteny information to identify relationships among different opsin classes. We used our phylogeny to identify significant genomic events that shaped opsin evolution and therefore the visual system of Limulus. We also describe the tissue expression patterns of the 18 opsins identified and show that transcripts encoding a number, including a peropsin, are present throughout the central nervous system. In addition to significantly extending our understanding of photosensitivity in Limulus and providing critical insight into the genomic evolution of horseshoe crab opsins, this work provides a valuable genomic resource for addressing myriad questions related to xiphosuran physiology and arthropod evolution.

Key words: Limulus polyphemus, xiphosuran, opsin, photoreceptors, evolution.

Introduction

The American horseshoe crab Limulus polyphemus (Linnaeus 1758) is one of four extant species of xiphosuran chelicerates, the sister group to arachnids (Regier et al. 2010; Edgecombe and Legg 2014). As such, studies of horseshoe crabs are key to understanding the origin of arachnids and the most recent
common ancestor of Euchelicerata. In addition, the Euchelicerata ancestor is a key node for better understanding the evolution of arthropods in general. *Limulus polyphemus*, hereafter referred to as *Limulus*, is the most studied of the extant horseshoe crabs, and investigations of its visual system have been central to understanding basic mechanisms of vision including phototransduction (e.g., Brown et al. 1984; Shin et al. 1993), light- and dark-adaptation (e.g., Lisman and Brown 1972; Behrens and Krebs 1976), visual information processing (Hartline et al. 1956) and the effects of circadian rhythms on visual function (Battelle 2013). Investigations of the *Limulus* visual system may also provide insights into the organization and function of visual systems in the most recent common ancestor of Arthropoda (Nilsson and Kelber 2007). Likewise, studies of genes encoding *Limulus* opsins, the protein component of the visual pigment, are central to understanding opsin evolution and diversification among chelicerates, a group in which knowledge of opsin proteins is limited.

Opsins have been classified into four major monophyletic groups: rhabdomeric or R-type opsins such as those found in the microvillar-rich photoreceptors in the eyes of arthropods, ciliary or C-type opsins such as those found in the ciliary rods and cones of vertebrates, cnidarians opsins or Cnidops, which appear unique to cnidarians, and retinal G-protein-coupled receptors (RGR)/Go-type or Group 4 opsins consisting of a mixed group of RGR, peropsins and neuropsins (reviewed in Porter et al. 2012). In previous studies, we determined that most photoreceptors in *Limulus* eyes express more than one R-type opsin.

*Limulus* has three different types of eyes: a pair of lateral compound, image-forming eyes called lateral eyes (LE), a pair of median ocelli called median eyes (ME) and three pair of larval eyes, lateral, median and ventral (fig. 1A). Among the larval eyes, the ventral larval eyes or ventral eyes (VE) have been studied most extensively. Five opsin genes—LpOps1–4, which encode nearly identical transcripts and therefore are considered a set (Dalal et al. 2003), and LpOps5—are co-expressed in LE retinular cells and giant photoreceptors in larval eyes (Katti et al. 2010), two opsins (LpOps5 and LpUVOps1) are co-expressed in small photoreceptors in larval eyes (Battelle et al. 2014), and three opsins (LpOps6, 7, and 8) are co-expressed in visible light sensitive photoreceptors in MEs (Battelle et al. 2015). We determined that in addition to being expressed in UV-sensitive ME photoreceptors, LpUVOps1 is expressed in LE eccentric cells (Battelle et al. 2014), a cell type originally thought to be a nonphotosensitive secondary cell (Waterman and Wiersma 1954). We showed further that a peropsin, LpPerOps1, is expressed in glia or pigment cells surrounding photoreceptors in each of the eyes (Battelle et al. 2015).

To aid in further studies of xiphosuran chelicerate visual systems in general, and of *Limulus* opsin genes in particular, we generated a high-quality genome assembly of *L. polyphemus*. Two previous studies have published genome sequences for *Limulus*, however the genome is large and resulting assemblies suffer from very low contiguity (Nossa et al. 2014; Kenny et al. 2016). The N50 values for these assemblies were all under 3 kb, which limits the types of analyses that can be performed. For example, large genes can be scattered across multiple scaffolds making it impossible to study gene structure, and it is impossible to study extended synteny on at least 50% of the genome. For these reasons, these sub-draft assemblies are not useful for an in-depth analysis of an extended gene family like the opsins. Using our high-quality genome assembly, we characterized the full repertoire of opsins in *Limulus*, conducted detailed phylogenetic analyses to classify each *Limulus* opsin, and used gene structure and synteny information to verify classifications and identify interesting genomic events that likely shaped the evolution of arthropod visual systems. Lastly, we provide detailed information on the tissue expression patterns for each of the 18 *Limulus* opsins, making this animal a pivotal resource for understanding the evolution of opsins and photosensitivity.

**Materials and Methods**

**Reagents**

Unless otherwise specified, reagents were purchased from Fisher Scientific (Pittsburgh, PA) or Sigma–Aldrich (St. Louis, MO).

**Genome Sequencing**

Genetic material used for sequencing was obtained during January 2008 from a single adult male (carapace length: ~16 cm) purchased from the Marine Biological Laboratory (Woods Hole, MA). Genomic DNA was prepared from limb muscle tissue, and source DNA (UCB-LP #5) is available from the lab of Nipam Patel at the University of California, Berkeley, CA.

We generated 18× sequence coverage (fragments, 3 and 8 kb) with reads generated on Roche 454 instrumentation. These combined sequence reads were assembled using the Newbler software (Roche Sequencing, Pleasanton, CA, a division of Hoffmann-La Roche LTD, Basel, Switzerland), and where possible, scaffold gaps were closed by mapping with 12× coverage of Illumina sequences and local gap assembly.

**Animals for Opsiin Experiments**

Adult *Limulus* were collected from the Indian River near Melbourne, FL (Latitude 28°42′ 31.46″ N; Longitude 80°44′ 53.06″ W). Young juveniles, between their first and second juvenile molts, were reared at the Whitney Lab following *in vitro* fertilization using eggs and sperm from adults also collected from the Indian River. Older juveniles, measuring 2.5–3.5 cm across the prosoma, were purchased from Pet Solutions (Bevercreek, OH). Adult animals were maintained...
in naturally running seawater held between 18 °C and 20 °C and fed shrimp twice a week. Juveniles were maintained in large containers of shallow, natural sea water over sandy bottoms. Twice a week, the seawater was changed and the juveniles were fed *Artemia*. All animals were maintained under natural illumination provided by a skylight in the aquarium room. Natural light intensities in the aquarium room were monitored continuously using a HOBO light data logger (Onset Computer Corporation, Pocasset, MA). They peaked midday at ~70,000 lux. The spectrum of light was also measured from 300–850 nm using an Ocean Optics USB4000 UV-visible spectrometer fitted with a 200-μm diameter UV-visible fiber (Ocean Optics, Dunedin, FL). No light with wavelengths <400 nm penetrated the skylight.

**Opsin Cloning**

The following opsins were cloned and characterized previously: LpOps1 and 2 (Smith et al. 1993), LpOps5 (Katti et al. 2010), LpUVOps1 (Battelle et al. 2014), and LpOps6, 7, 8 and LpPerOps1 (Battelle et al. 2015). Previous studies also identified two additional LpOps1-like genes called LpOps3 and 4 (Dalai et al. 2003) and a second peropsin gene called LpPerOps2 (Battelle et al. 2015).

**RNA Isolations and cDNA Preparation**

We used RNeasy (Qiagen, Valencia, CA) to isolate RNA from the following tissues from older juveniles: brain, tail and the synganglion pooled with all segmental ganglia. We also used RNeasy to isolate RNA from adult MEs and VEs. We used RNAzol to isolate RNA from adult LE, brain, synganglia and all segmental ganglia pooled together. To prepare RNA from large juveniles, we pooled tissues from three or four animals. To prepare RNA from adults, we used one LE, eight MEs, eight VEs, one or two brains, one synganglion and all segmental ganglia pooled from a single animal. In some instances, we removed a portion of the brain anterior to the optic ganglia to reduce contamination by ventral photoreceptors attached to the brain. RNA was reverse transcribed with SuperScript III-First Strand Synthesis System for RT-PCR (Life Technologies, Grand Island, NY). The cDNA library was prepared from RNA isolated from the entire CNS of young juveniles (Katti et al. 2010). All animals were sacrificed and RNA was extracted in the morning in the light.

---

**Fig. 1.**—Schematics showing the locations of *Limulus* eyes and the structure of its CNS. (A) Dorsal view of an adult animal showing the locations of its eyes. The upper box is an enlargement showing the median ocelli and the fused median larval eyes between them. The box on the left shows an enlargement of a lateral compound eye and the location of the lateral larval eye at its posterior edge. The cut-away in the center shows the locations of the brain and ventral optic nerves projecting from the brain to the end organ. The synganglion posterior to the brain is also shown. (B) Dorsal and (C) ventral view of the CNS of a juvenile animal measuring ~2–2.5 cm across the prosoma. BR, brain; CB, central body; CP, corpora pedunculata; L, lamina; LON, lateral optic nerve; M, medulla; MON, median optic nerve; ON, ocellar neuropile; SG, segmental ganglia (abdominal ganglia); SY, synganglion (circumesophageal ring); VON, ventral optic nerve. Scale bar, 1 mm.
In the present study, we identified sequences encoding portions of five additional presumptive R-type opsins in a TBLASTN search of our *Limulus* genome assembly (GenBank accession: GCA_000517525.1) with LpOps1 (Accession number AAA02499) as query. As we detail below, we used PCR to amplify putative opsins from cDNA prepared from various *Limulus* tissues and cloned partial or full-length open reading frames (ORF) into pGem-T (Promega Corp. Madison, WI). The primers we used are listed in supplementary table S1, Supplementary Material online.

We identified LpUVOps2 as a 650-base pair (bp) opsin-like genomic sequence and amplified cDNA encoded by this sequence from adult LE and CNS using primers UVOps2F1 and UVOps2R1. We obtained its full-length ORF using a RACE (Rapid Amplification of cDNA Ends) strategy (Katti et al. 2010) with the cDNA library from young juvenile CNS as template (5' RACE primers: UVOps2R1 and CAP followed by UVOps2R2 and CAP; 3' RACE primers UVOps2F1 and TRSALu4 followed by UVOps2F2 and Lu4NS). We amplified the full-length ORF from adult brain cDNA with primers UVOpsF6 and UVOpsR4 (LpUVOps2 Accession number KU40433).

We identified LpOps9 as a 731-bp opsin-like genomic sequence and amplified cDNA encoded by this sequence from adult LE, VE, ME, brain, segmental ganglia and juvenile tail using primers F1 and R1. We predicted the full-length ORF from the genome assembly and amplified it from adult brain cDNA with primers F2 and R2 (LpOps9 Accession number KU40434). We identified LpOps10 as a 665-bp opsin-like genomic sequence and amplified cDNA encoded by this sequence from the young juvenile CNS cDNA library and juvenile tail cDNA using primers F1 and R1. We extended the sequence toward the 5' and 3' ends with primers F6 and R5 designed based on the genomic assembly and cloned the resulting 1008-bp piece (LpOps10 Accession number KU40435). We were unable to obtain a full-length clone either by RACE or by PCR using primers designed based on the genomic assembly.

We identified LpArthOps1 and LpArthOps2 as 750- and 657-bp opsin-like genomic sequences, respectively. We amplified cDNA encoded by the LpArthOps1 genomic fragment with primers F1 and R3 from adult brain and segmental ganglia cDNA. We obtained its full-length ORF with the RACE strategy described above using the cDNA library from young juvenile CNS as template. Gene-specific primers for the 5' RACE were R5 followed by R4; for the 3' RACE they were F1 followed by F3. We amplified the full-length ORF with primers F4 and R6 using the cDNA library from young juvenile CNS as template (LpArthOps1 Accession number KU40431). We amplified cDNA encoded by the LpArthOps2 genomic fragment with primers F1 and R3 from young juvenile CNS cDNA. We predicted its full-length ORF from the genome assembly based on its homology with LpArthOps1, and we amplified its full-length ORF from the young juvenile CNS cDNA library with primers F4 and R6 (LpArthOps2 Accession number KU40432).

We additionally found portions of two presumptive *Limulus* C-type opsin genes, LpCOPS1 (585 bp) and LpCOPS2 (592 bp), with a TBLASTN search of the *Limulus* genome assembly using a C-type opsin amino acid sequence from a spider (Accession number CCP46950) (Eriksson et al. 2013) as query. We amplified cDNAs encoded by these genomic sequences from adult brain and segmental ganglia cDNA with LpCOPS1 primers F1 and R1 and LpCOPS2 primers F2 and R2. Using adult brain cDNA as template, we extended the LpCOPS1 sequence toward the 3' end with primers F6 and R10 designed based on the genomic assembly and obtained an 831-bp sequence. We extended LpCOPS2 toward the 5' end with primers F12 and R4 and obtained a 740-bp sequence. We were unable to obtain full-length sequences for either C-type opsin by RACE or by PCR using primers based on gene predictions. Accession numbers of the partial sequences of LpCOPS1 and 2 are KU40436 and KU40437, respectively.

**Opsin Gene Phylogenetic Analysis**

In order to place the identified *Limulus* opsins within the most current understanding of opsin evolution and classification, we reconstructed a phylogeny using recent large opsin data sets (Feuda et al. 2012, 2014; Porter et al. 2012; Henze and Oakley 2015) as well as more recently published arthropod opsin data (e.g., Eriksson et al. 2013; Hering and Meyer 2014; Hwang et al. 2014). We added to these data sets several non- *Limulus* chelicerate opsin sequences including those from the spider mite *Tetranychus urticae* (Grbić et al. 2011) and those we mined from the genomes of the scorpion *Mesobuthus martensi* (Cao et al. 2013) (supplementary table S2, Supplementary Material online) and two other horseshoe crabs, *Carcinoscorpius rotundicauda* and *Tachypleus tridentatus* (Kenny et al. 2016) (supplementary table S3, Supplementary Material online), using *Limulus* opsins as queries for TBLASTN.

Two phylogenies were reconstructed for understanding eucharistic, and specifically *Limulus*, opsin evolution: one large data set of 743 genomic and expressed sequences representing the known evolutionary diversity of opsin proteins and one small data set consisting of all known *Limulus* opsins. Both data sets consisted of amino acid sequences aligned with MAFFT (Katoh et al. 2002; Katoh and Standley 2013). The resulting alignments were used to estimate phylogenetic relationships and node confidence as bootstrap values using RAxML (Stamatakis 2006, 2014; Liu et al. 2012 Stamatakis et al. 2008; Pattengale et al. 2009) with a GTR+G model of evolution (Feuda et al. 2014) as implemented in CIPRES (Miller et al. 2010). Both phylogenies were rooted using related GPCR melatonin receptors and/or *Trichoplax adhaerens* sequences as out groups as outlined in Feuda et al. (2014). For both phylogenies, opsin amino acid sequence alignments,
sequence database information, and newick tree files have been deposited on DRYAD digital repository (doi:10.5061/dryad.k43t2).

Intron positions and phases of LpOps1–4 were determined previously (Dalal et al. 2003). We identified intron positions and phases of other opsins from TBLASTN analyses of the Limulus genome assembly and from well-assembled genomes of other species.

There were two cases where the topology of our tree suggested extraordinary evolutionary findings, whereas an alternative topology would lead to a simpler explanation. The simpler alternative hypotheses were: (1) a monophyletic LpOps6, LpOps7 and LpOps8, which are uniquely expressed in ME and (2) a monophyletic LpOps9, LpOps10, LpUVOps1, LpUVOps2, LpArthOps1 and LpArthrop2, which all share a very similar intron/exon structure. To statistically test these hypotheses, we performed a Swofford–Olsen–Waddell–Hillis (SOWH) test (Swofford et al. 1996) for each scenario. We used the program SOWHAT (version 0.35) (Church et al. 2015) to carry out these analyses on an alignment consisting of the Limulus opsin proteins. We specified the PROTGAMMM AWAG model as implemented in RAxML (version 8.1.21) (Stamatakis 2006) using the default 1000 replicates for both tests.

Distributions of Opsin Transcripts and Proteins

RT-PCR

We used RT-PCR to probe for transcripts encoding each opsin in cDNAs from the following tissues: ME, VE, LE, brain, synganglion and pooled segmental ganglia from adult animals, and brain, tails and the synganglion pooled with all segmental ganglia from large juveniles. The primers we used in most screens (supplementary table S4, Supplementary Material online) were designed to amplify across an intron and eliminate the possibility of amplifying genomic DNA. The exceptions were screens for LpOps6 and 7, which lack introns. In all screens, we assayed cDNAs prepared from at least two different tissue collections and verified the identity of each PCR product by sequencing.

In Situ Hybridization

We prepared sense and antisense digoxigenin-labeled RNA probes from the full-length coding regions of LpOps1, LpOps5, LpUVOps1, and LpPerOps1 as described previously (Katti et al. 2010; Battelle et al. 2014, 2015). Because LpOps1, 2, 3 and 4 transcripts are nearly identical, the probe directed against LpOps1 will detect all four transcripts. We also prepared digoxigenin-labeled sense and antisense RNA probes from full-length clones of LpUVOps2 and LpArthOps1, the 1008-bp fragment of LpOps10, and 585- and 592-bp fragments of LpC-Ops1 and LpC-Ops2, respectively. We used all probes at a final concentration of 1 μg/μl. We applied probes to whole mounts of ventral larval eyes dissected from adults and CNS tissues (brain, synganglion and segmental ganglia) dissected from large juveniles that had been fixed and processed for in situ hybridization as previously described (Jezzini et al. 2005) except that we exposed tissues to probes for 72 h at 65 °C and used the color development protocol described by (Seaver and Kaneshige 2006). The times for color development ranged from a few hours to 7 days. We photographed whole mounts with a Zeiss Discovery VS stereo microscope. Fixed, frozen sections of LE and ME were probed for opsin transcripts as described previously (Battelle et al. 2015).

Immunocytochemistry

We fixed, processed and cut serial frozen sections of CNS tissue from large juveniles as described previously (Katti et al. 2010). We sectioned brain, synganglia and segmental ganglia separately, and immunostained sections as detailed previously for myosin III (LpMyoIII), LpOps1–2 (Battelle et al. 2001); LpOps5 (Katti et al. 2010) and LpPerOps1 (Battelle et al. 2015). The specificity of each antibody was verified in the studies cited above. After immunostaining, we incubated some sections with DAPI to visualize DNA. We collected fluorescent images using a confocal microscope as described previously (Battelle et al. 2015).

Results

Genome Sequencing

We sequenced the genome of a single adult Limulus male using a combined approach of Roche 454 and Illumina sequencing. We used K-mer frequencies to estimate the genome size to be ~1.5 gigabases (Gb). The assembly is comprised of 286,792 scaffolds with an N50 scaffold length of 238 kb and an N50 contig length of 11.4 kb (table 1). The assembled coverage is 18x, and the assembly spans 1.8 Gb. We removed all contaminating sequences from the assembly, trimmed vectors (X), and ambiguous bases (N). Additionally, shorter contigs (<200 bp) were removed prior to public release. In table 1, we also compare our assembly of the Limulus genome with assemblies reported previously for Limulus and two other extant horseshoe crabs, C. rotundicauda and T. tridentatus (Nossa et al. 2014; Kenny et al. 2016). The annotation for Limulus was generated by the National Center for Biotechnology Information (NCBI). The analysis identified 22,129 genes and pseudogenes, 2066 transcripts, and a total of 23,287 coding DNA sequences. The mean length of all genes is 29,383 bp.

Identification of Limulus Opsin Genes

We identified 18 Limulus opsin genes by BLASTing R-type and C-type opsin sequences against our Limulus genome assembly. The predicted proteins encoded by these genes are clearly opsins (supplementary fig. S1, Supplementary Material online) were designed to amplify across an intron and eliminate the possibility of amplifying genomic DNA. The exceptions were screens for LpOps6 and 7, which lack introns. In all screens, we assayed cDNAs prepared from at least two different tissue collections and verified the identity of each PCR product by sequencing.
online). Each has seven predicted transmembrane domains, a predicted conserved lysine in helix VII that is critical for Schiff base binding of the chromophore, suggesting that each can form a photopigment, and acidic amino acids (glutamic acid/aspartic acid) at sites 83 and 181 (bovine rhodopsin numbering), which are potential sites for the Schiff base counter-ion in some R-type opsins (Porter et al. 2012). The sequences also have other motifs specific to different opsin classes (see below).

**Relationship of Limulus Opsins to One Another and to Opsins of Other Species**

To place the Limulus opsin genes within known opsin sequence diversity, we reconstructed a maximum-likelihood phylogeny that included 743 sequences representing known opsin and taxonomic diversity (fig. 2). Included in this tree are partial sequences of five opsins recovered from the scorpion genome (supplementary table S2, Supplementary Material online), only two of which (Mmops1 and Mmops2) were described previously (Cao et al. 2013). Mmops3 described by Cao et al. (2013) was not included because it lacks a lysine in the chromophore binding pocket and therefore is probably not an opsin. We also include partial sequences of 14 opsins recovered from the *T. tridentatus* genome and 15 from the *C. rotundicauda* genome (supplementary table S3, Supplementary Material online).

We placed *Limulus* opsins within three of the four major opsin groups: R-type, C-type and RGR/Go-type, also referred to as Group 4 opsins (Porter et al. 2012). Among the *Limulus* R-type opsin genes, we identified the following: seven long-wavelength sensitive (LWS) opsins (LpOps1–4, 6, 7 and 8), one middle-wavelength sensitive (MWS) opsin (LpOps5), one ultraviolet sensitive (UVS) opsin (LpUVOps1), three that are most closely related to pancrustacean UV7 opsins (LpOps9, 10 and UVOps2), and two arthropsins (LpArthOps1 and 2). Each of these opsins contains an eight-amino acid indel found in many arthropod opsins that lengthens cytoplasmic loop III (Porter et al. 2007). The sequence of this indel, which is highly conserved in crustacean R-type opsins, is also highly conserved in many *Limulus* R-type opsins (LpOps1–4, 5, 6, 8, 10 and UVOps1), moderately conserved in LpOps7, but poorly conserved in LpUVOps2, LpOps9 and the arthropins (supplementary fig. S1, Supplementary Material online). The amino acid triplet characteristic of R-type opsins that activate the G-protein Gαq/11 (HPK/R) is conserved in all *Limulus* opsins within the R-type opsin group except for the two arthropsins (supplementary fig. S1, Supplementary Material online). This raises questions about the identity of the down-stream targets of arthropsins.

In addition to the R-type opsins, we identified two C-type opsins (LpC Ops1 and 2) and two peropsins (LpPerOps1 and 2) in the *Limulus* genome. LpC Ops1 has the amino-acid triplet NPQ, which is similar to and aligns with the sequence NKQ found in vertebrate C-type opsins thought critical for activating Gq. The sequence of LpC Ops2 in this region is not yet known. In LpPerOps1 and 2 the sequence at this site is NPR (supplementary fig. S1, Supplementary Material online), which is similar to and aligns with the sequence NPK in spider peropsins (Nagata et al. 2010; Eriksson et al. 2013) and the sequence HKK and NKK in mouse and human peropsins, respectively (Sun et al. 1997). The G-protein activated by peropsins, if any, is unknown.

In our analyses, most of the *Limulus* opsins form closely related clades with opsins from the other chelicerates: spider, tick, mite and scorpion. Furthermore, in the other two extant horseshoe crabs *T. tridentatus* and *C. rotundicauda*, we found the same diversity of opsins as we found in *Limulus*, including homologues of LpOps5, which cluster among crustacean MWS opsins (fig. 2). We assembled only one complete LpOps1-like sequence from the genomes of *T. tridentatus* and *C. rotundicauda*, but in the genomes of both of these species, we found nearly identical LpOps1-like sequences on multiple short contigs suggesting that they also have multiple Ops1-like genes.

A number of *Limulus* opsin genes appear to be paralogous groups: LpOps1–4, LpOps6 and 7, LpUVOps2 and LpOps9, LpArthopsin1 and 2, LpPerOps1 and 2, and LpC Ops1 and 2

<table>
<thead>
<tr>
<th>Species</th>
<th>Current study</th>
<th>Nossa et al. (2014)</th>
<th>Kenny et al. (2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequence length (bp)</strong></td>
<td>L. polyphemus</td>
<td>L. polyphemus</td>
<td>L. polyphemus</td>
</tr>
<tr>
<td></td>
<td>1,828,256,766</td>
<td>1,229,280,963</td>
<td>1,446,611,838</td>
</tr>
<tr>
<td>Assembly gap length (bp)</td>
<td>122,485,139</td>
<td>456,658</td>
<td>NA</td>
</tr>
<tr>
<td>Scaffold number</td>
<td>286,792</td>
<td>896,522</td>
<td>NA</td>
</tr>
<tr>
<td>Scaffold N50 (bp)</td>
<td>254,089</td>
<td>2,929</td>
<td>NA</td>
</tr>
<tr>
<td>Contig number</td>
<td>469,509</td>
<td>6,614,434</td>
<td>4,214,715</td>
</tr>
<tr>
<td>Contig N50</td>
<td>11,441</td>
<td>418</td>
<td>466</td>
</tr>
</tbody>
</table>

**NOTE.**—bp, base pairs; NA, not applicable.
**Fig. 2.**—Opsin phylogeny. Maximum-likelihood tree of 743 genomic and expressed opsin sequences illustrating the four major evolutionary clades of opsins (C-type, R-type, Cnidarian, and Group 4), divided into subclades by major taxonomic groups (see key for color codes). Panarthropod groups are expanded to highlight the relationships among *Limulus polyphemus* (in bold) and arachnid and other xiphosuran opsin sequences. For the R-type clade containing the known panarthropod opsin genes, the spectral clades have been indicated as long wavelength sensitive (LWS), middle wavelength sensitive (MWS), and short wavelength sensitive (SWS). Black circles on branches indicate nodal support >80%.
Each set has an identical gene structure (intron location and phase) and encodes proteins that share 50% or greater sequence identity (supplementary fig. S2, Supplementary Material online). Most sets are encoded on different genomic scaffolds, but the genes encoding LpOps1–4 are on the same scaffold, and the proteins these genes encode are 99% identical to one another suggesting they are the result of tandem duplication events. LpOps1 and 2 transcripts can be distinguished unambiguously from one another by the sequences of their 3₀-untranslated regions (Smith et al. 1993), but LpOps1, 3 and 4 genes can be distinguished from one another only by their intron sequences (Dalal et al. 2003).

It has recently been proposed that at least one whole-genome duplication event occurred in the stem ancestor of modern-day horseshoe crabs (Nossa et al. 2014; Kenny et al. 2016). This suggestion is based on the identity of pairs of loci with large numbers of shared paralogous gene pairs on different chromosomes including the presence of duplicated Hox and ParaHox clusters in *Limulus*, *T. tridentatus* and *C. rotundicauda*. To test if the paralogous opsin pairs on distinct scaffolds were consistent with the proposed ancient whole-genome duplication, we BLASTed these scaffolds against the human reference protein set and looked for other genes that might be shared between these scaffolds. We found examples of top BLAST hits on scaffolds that encode paralogous copies of the arthropins, the peropsins and LpOps7 and LpOps6 and 8 (supplementary fig. S3, Supplementary Material online). LpOps6 and 8 are located in tandem on the same scaffold. These observations are consistent with the proposed whole-genome duplication. But LpOps5, 8, 10, and UVOps1 do not have paralogs. Thus, if a whole-genome duplication occurred, it must have been followed by significant opsin gene loss.

When we compared *Limulus* opsin genes with those we recovered from the genomes of two other extant horseshoe crabs, *T. tridentatus* and *C. rotundicauda*, we found the same complement of paralogous opsin pairs and nonparalogous opsins in all three species (fig. 2). This finding is consistent
with the proposed whole-genome duplication occurring early in the xiphosuran lineage (Kenny et al. 2016) and suggests that opsin gene loss also occurred early in the lineage. The argument for a whole-genome duplication event early in the xiphosuran lineage would have been strengthened had we discovered synteny between paralogous opsin pairs and other genes in the genome assemblies of *T. tridentatus* and *C. rotundicauda*. Unfortunately, we were unable to perform these analyses because the genome assemblies of *T. tridentatus* and *C. rotundicauda* are too fragmented (table 1).

We examined gene structure to explore further relationships among *Limulus* opsins and between *Limulus* opsins and opsins from other species. We were particularly interested in the relationship between LpOps6 and the paralogs LpOps6 and 7. All three are uniquely expressed in MEs (see below) and are co-expressed in one population of ME photoreceptors (Battelle et al. 2015). Furthermore, LpOps6 and 7 are intronless paralogs, and as was mentioned above, the LpOps6 and 8 genes are located on the same scaffold within 5.5 kb of one another. While the linkage of LpOps6 and 8 is consistent with the occurrence of a recent tandem duplication event, our phylogenetic analyses (figs. 2 and 3A) suggest that LpOps8 is more closely related to LpOps1–4 than it is to LpOps6 and 7. To be sure that LpOps6, 7, and 8 are not a monophyletic clade, we ran a Swofford–Olsen–Waddell–Hillis (SOWH) test, which rejected this alternative hypothesis ($p<0.001$). Our analysis of gene structure, which shows that intron two in LpOps8 aligns with intron one of LpOps1–4, provides additional support for a close relationship between LpOps1–4 and 8 (figs. 2 and 3B).

Using our phylogeny, we reconstructed an evolutionary scenario explaining the origins of the *Limulus* LWS genes (fig. 4). The scenario is consistent with an ancient whole-genome duplication event occurring in the Xiphosura stem lineage. Furthermore, the scenario suggests that the synteny between the median-eye specific LpOps6 and 8 opsins has been maintained for a very long time and therefore may have a strong functional significance.

We found that opsins from different clades have the same gene structure. For example introns 4 and 5 in LpPerOps1, a RGR/Go-type opsin, match in position and phase the introns in LWS LpOps1–4. We also found that opsins LpUVOps1 and 2, LpOps9 and 10, and LpArthops1 and 2 each has two introns with identical positions and phases. Our maximum-likelihood trees (figs. 2 and 3A) place these opsins in three distinct R-type opsin clades: (1) chelicerate UV7 opsins within the larger pancrustacean short-wavelength sensitive (SWS) opsin clade, (2) chelicerate UVOps2, Ops9 and Ops10 within the pancrustacean UV7 opsin clade, and (3) the more distantly related arthropins, containing both crustacean and chelicerate sequences. Their identical gene structure suggested they might be a monophyletic group, but based on the results of
a SOWH test, we rejected this alternative hypothesis in favor of the relationship shown in the maximum-likelihood tree \( (p \leq 0.001) \).

Our observation that different clades of Limulus R-type opsins have the same gene structure prompted us to examine whether opsins from other species also have this structure. To test this, we used TBLASTN to identify R-type opsin genes in selected genomes using LpOps10 as query. We recovered R-type opsin genes with the same two introns from representative species of arthropods, annelids, echinoderms and mammals (supplementary fig. S4, Supplementary Material online). We found that the arthropsins of the water flea Daphnia pulex also have the same two introns (not shown). By contrast, the intron structure of the LWS Limulus opsins, which are most closely related to the SWS Limulus opsins based on sequence homology and phylogenetic analyses, do not share these two introns.

We found that intron-exon junctions in other Limulus opsins are conserved in distantly related lineages. The three introns in Limulus C-type opsins are strictly conserved with introns 1, 3 and 4 in bovine rhodopsin (supplementary fig. S5, Supplementary Material online), although the donor sequence of the first intron of LpCops2 is AC instead of the canonical GT (Burset et al. 2000) identified in bovine rhodopsin (Nathans and Hogness 1983) and in all other introns of Limulus opsins (this study, not shown). Introns 1, 3 and 5 of LpPerOps1 are strictly conserved with peropsin genes of vertebrates, hemichordates and several other invertebrates (Albalat 2012). Intron 1 in LWS LpOps1–4 and 8 is strictly conserved in insect LWS opsins and the second intron in LWS LpOps1–4 matches in position and phase an intron in some insect LWS opsins (supplementary fig. S6, Supplementary Material online). Although the introns of LpOps5 are unique among Limulus opsins (fig. 3B, supplementary fig. S1, Supplementary Material online), intron 2 of LpOps5 matches the position and phase of intron 4 in each of the D. pulex opsins most closely related to LpOps5 within the larger MWS clade (fig. 2; supplementary fig. S7, Supplementary Material online) but does not align with any introns in D. pulex opsins in the second subgroup within the MWS clade (not shown).

**Opsin Expression**

**Organization of the Limulus Visual and Nervous Systems**

The schematics of the Limulus visual and central nervous systems in fig. 1 will orient readers to structures we assayed for opsin expression. In fig. 1A, we show the locations of the three types of Limulus eyes, and in the central cut-away, we show the location the brain. Also, in the central cut-away, we show the optic nerves of the VEs, which in adult animals, project anteriorly from the brain and terminate in a pair of end organs attached to a specialized region on the ventral cuticle. Each end organ typically contains a large cluster of photoreceptor cell bodies. In newly hatched animals and juveniles, ventral photoreceptor cell bodies lie close to the anterior brain, and even in adults, some ventral photoreceptor cell bodies remain on the brain (Chamberlain and Wyse 1986). On the brain’s dorsal side (fig. 1B), central projections from the lateral, median and ventral optic nerves can be seen, as well as the lamina (first optic ganglia), medulla (second optic ganglia) and central body. The corpora pedunculata, also called mushroom bodies, are on the ventral side of the brain as are most neuronal cell bodies of the synganglion and segmental ganglia (fig. 1C).

**Tissue Distribution Assayed with RT-PCR**

We screened for transcripts encoded by each of the 18 opsin genes in cDNA prepared from adult eyes and in a cDNA library prepared from young juveniles. Except for LpPerOps2, we found that each of the opsin genes identified in the genome is expressed in one or more of these tissues, with one qualification. Because transcripts encoded by the LpOps1, 3, and 4 genes are nearly identical (Dalal et al. 2003), we have no direct evidence that all three genes are expressed.

Nine Limulus opsins and their expression patterns in eyes were described in detail previously: LpOps1 and 2 (Smith et al. 1993; Dalal et al. 2003), LpOps5, (Katti et al. 2010), LpOps6, 7, 8 and LpPerOps1 and 2 (Battelle et al. 2015), and LpUVOps1 (Battelle et al. 2014). In the current study, we screened for transcripts encoded by the newly identified opsin genes (LpUVOps2, LpOps9 and 10; LpArthops1 and 2; and LpCops1 and 2) in cDNA from adult eyes, the CNS of adults and the CNS and tail of older juveniles. We also searched for transcripts encoding the previously identified opsins in cDNA from the CNS of adults and the CNS and tail of older juveniles. Because all larval eyes contain the same two classes of photoreceptors (Harzsch et al. 2006; Battelle et al. 2014) we consider results from assays of adult VEs as representative of all three types of larval eyes. Figure 5 summarizes results obtained from adult eyes and CNS and from juvenile tails. The distribution of opsin transcripts in the CNS of the older juveniles, which were also used in the in situ assays, was the same as that observed in adults. Supplementary Figure S8, Supplementary Material online shows sample results from the PCR reactions used to generate the results in fig. 5.

We detected transcripts of all opsins in one or more of the eyes, CNS or tail, except for LpPerOps2, which is not expressed in any of the tissues we assayed. LpArthops2 transcripts were not detected in any tissues we assayed from older juveniles or adults, but we obtained a full length clone of this opsin from the CNS of young juveniles indicating it may be expressed early in development. We detected most opsin transcripts in multiple tissues. Exceptions were transcripts encoding LpOps6, 7 and 8, which we detected only in MEs, and
**FIG. 5.** —Distribution of opsin transcripts and protein in *Limulus* eyes and nervous system. (A) Schematic of *Limulus* showing the locations of tissues tested for opsin transcripts by PCR and in situ hybridization and for opsin protein by immunocytochemistry. (B) (Left) Phylogenetic tree of *Limulus* opsins is modified from that shown in fig. 3. LpPerOps2 is not included here because transcripts encoding this opsin were not detected in any tissues assayed. LpArthOps2 is included because its transcripts were detected in the CNS of young juveniles. (B) (Right) Tissue distribution of the opsin shown at left. (Top) Tissues assayed. Tissues in which photosensitive cells have been detected by electrophysiology are indicated with an asterisk (*). (See Discussion for details). Solid blue box, transcript detected by PCR and in situ hybridization/immunocytochemistry. Blue-white box, transcript detected by PCR but not by in situ hybridization/immunocytochemistry. Blue-gray box, transcript detected by PCR and not tested by in situ hybridization/immunocytochemistry. Solid white box, transcript not detected by PCR and therefore not tested with in situ hybridization. The distribution of some opsins in eyes was determined previously. (a) Smith et al. 1993, (b) Dalal et al. 2003, (c) Katti et al. 2010, (d) Battelle et al. 2015, and (e) Battelle et al. 2014. ME, median eye; VE, ventral eye; LE, lateral eye; BR, brain; SY, synganglion; SG, segmental ganglia; TL, tail.
LpOps10, which we detected only in the tail. We found that three of the sets of opsin paralogs have the same expression pattern (i.e., LpOps1–4; LpOps6 and 7; LpCOps1 and 2) and three do not (i.e., LpOps9 and LpUVOps2; LpPerOps1 and 2; LpArthrop1 and 2). We detected LpPerOps1 in all tissues assayed but LpPerOps2 in none. Similarly, we detected LpArthrop1 throughout the CNS, in the tail and VEs, but we found LpArthrop2 in none of these tissues in older juvenile or adult animals.

**Cellular Distribution Assayed with In Situ Hybridization and Immunocytochemistry**

We used in situ hybridization assays to examine the cellular distribution of opsin transcripts in adult eyes and in the CNS of older juveniles. We assayed for opsin proteins in these same tissues using immunocytochemistry and specific antibodies directed against LpOps1–4, 5, 6, UVOps1 and PerOps1. We were unable to apply these techniques to the tail because of the unique challenges of doing morphology on the tail.

**Eyes**

As was described above, our PCR screens detected LpOps9 transcripts in each eye type and LpUVOps2 and LpArthrop1 transcripts in VEs (fig. 5). However, our in situ hybridization assays for these transcripts in eye tissues consistently produced negative results. By contrast, using identical in situ hybridization protocols, we detected LpOps1, 5 and LpUVOps1 in LEs, LpOps 6, 7, 8 and LpUVOps1 in MEs and LpOps 1, 5 and LpUVOps1 in VEs (Battelle et al. 2014, 2015).

**CNS**

We consistently detected LpOps1–4 transcripts in processes of LE photoreceptors as they enter the brain and terminate in the lamina, in VE photoreceptor cell bodies located close to the brain and in VE processes that project to and terminate in the medulla (fig. 6A and 8). We also detected LpOps5 and LpUVOps1 transcripts in VE photoreceptor cell bodies and their processes (fig. 6C and D). We did not detect LpOps1–4, 5 or LpUVOps1 transcripts in cell bodies or processes elsewhere in the brain, and although our PCR screen revealed LpOps1–4 and 5 transcripts in the synganglion and abdominal ganglia (fig. 5), we were unable to detect transcripts in these tissues with in situ hybridization assays even after 7 days of development. Using antibodies that detect LpOps1–4, 5, and UVOps1 proteins in rhabdoms of photoreceptors in eyes (Battelle et al. 2001, 2014; Katti et al. 2010), we detected these opsins in ventral photoreceptor cell bodies on the brain, but not in LE or VE processes or elsewhere in the brain (not shown). We also were unable to detect these opsins in sections of the synganglion or segmental ganglia. Our in situ hybridization assays for LpUVOps2, LpOps9, LpArthrop1 and LpCOps1 and 2 transcripts in whole mounts of brain and ventral nerve cord of older juveniles also produced negative results.

By contrast, our in situ hybridization assays revealed LpPerOps1 transcripts throughout the CNS (fig. 7A–F). In the brain, the antisense probe targeting LpPerOps1 labeled ventral optic nerves most intensely (fig. 6A and 8), which is consistent with our finding of LpPerOps1 immunoreactivity in glia surrounding ventral photoreceptors (fig. 7G and Battelle et al. 2015). We detected no LpPerOps1 transcripts in the corpora pedunculata (fig. 7B), but transcripts were consistently detected in cells at the periphery of the lateral optic nerves and in what appear to be fibers within the central body (fig. 7A). In the synganglion, transcript was associated with neuronal clusters located between large nerve bundles projecting to the periphery. These cell clusters were most evident when viewed from the dorsal side (fig. 6C). In segmental ganglia, transcript was consistently associated with two or three bilateral neuronal clusters in each ganglion. When the brain, synganglion and segmental ganglia were immunostained for LpPerOps1 protein, we detected LpPerOps1 immunoreactivity surrounding neurons indicating LpPerOps1 protein is present in these regions (fig. 7G–J).

**Discussion**

In this study, we sequenced and assembled the genome of the American horseshoe crab L. polyphemus. We used these data to identify and phylogenetically classify 18 opsin genes in this genome into three of the four currently recognized major opsin groups and gain insights into a number of key events in arthropod opsin evolution. We further showed that gene structure supports the placement of the opsins within these major clades and provides new insights into the relationship of the major clades to one another. In our studies of opsin expression, we detected transcripts for all of these opsins by RT-PCR in the eyes, CNS or tail of adults or older juveniles except for LpArthrop2, which was detected only in the CNS of young juveniles and LpPerOps2, which was not detected in any tissues we assayed. We also showed that LpPerOps1 is expressed in glia throughout the CNS. We were unable to identify opsin-expressing neurons in the CNS by in situ hybridization or immunocytochemistry suggesting that opsin transcripts levels are low (or absent) in neurons, which raises questions about their functional significance. However, as is discussed below, electrophysiological studies have shown that photosensitive cells are present in Limulus CNS and tail. Our results provide a list of opsin candidates that may contribute to this extracellular photosensitivity.

**Arthropod Opm Evolution**

Our data point to key events in the evolution of arthropod opsins. The large number of opsins in horseshoe crabs—18 in Limulus, and at least 14 and 15 in T. tridentatus and C. rudenticauda, respectively—compared with the number so far
identified in other chelicerates—six in transcriptomes from the jumping spider *C. salei* (Eriksson et al. 2013), six in the spider *S. mimosarum*, and five that we recovered from the genome of the scorpion *M. martensii*)—is consistent with a whole-genome duplication event early in the xiphosuran lineage (Nossa et al. 2014; Kenny et al. 2016). Similar patterns are not present in scorpions or spiders, which suggests that the event occurred after Xiphosura diverged from the rest of Euchelicerata. As the same complement of paralogous opsin pairs and nonparalogous opsins are present in all three horseshoe crab species examined, we suggest that significant opsin gene loss also occurred in the stem of the xiphosuran lineage.

Several opsins appear to have evolved only in xiphosurans. For example, LpOps6, 7, and 8, which are ME specific in *Limulus*, appear to have radiated from an ancestral LWS opsin after Xiphosura diverged from the rest of Euchelicerata. We also identified opsins in xiphosurans that appear to have been lost in other euchelicerates. We found
**Fig. 7.**—Distribution of LpPerOps1 transcripts and protein. LpPerOps1 transcripts and proteins were detected in the brain, synganglion and segmental ganglia of older juveniles. (A–F) Representative whole-mount of the CNS from an older juvenile incubated with an antisense probe targeting LpPerOps1 transcripts. Dorsal views (A, C, and E) and ventral views (B, D, and F) of the brain (A and B), synganglion (C and D) and segmental ganglia (E and F). In the brain (A and B), LpPerOps1 transcripts were detected in the ventral optic nerve (VON), at the periphery of the lateral optic nerve (LON), in fibers that may be in the central body (CB), which is visible on the dorsal side of the brain. The locations of the second optic ganglia or medulla (M) are indicated. Transcripts were not detected in the corpora pedunculata (CP) on the brain’s ventral side. Transcripts in the synganglion (C and D) were associated with cell clusters located between the large nerve roots projecting to the periphery. In segmental ganglia (E and F), transcripts were typically associated with two bilateral cell clusters in each ganglion (arrow heads). Scale bars, 1 mm. (G–J) Fixed, frozen sections of cell clusters from different regions of the juvenile CNS that were immunostained for LpPerOps1 (Green) and incubated with DAPI to reveal nuclei (Blue). (G) A small cluster of giant ventral photoreceptors on the brain. Ventral photoreceptor cell bodies, identified with LpMyoIII-immunoreactivity (red) (LpMyoIII-ir), a marker for photoreceptors in each of the eyes (Battelle et al. 2001), are surrounded by LpPerOps1-immunoreactive glia, as was described previously (Battelle et al. 2015). (H) Immunostained cell clusters from the synganglion. The periphery of the cluster is outlined. (I and J) Cell clusters from different segmental ganglia, as indicated. In each cell cluster examined, LpPerOps1-immunoreactivity (LpPerOps1-ir) surrounded neurons and was not uniform throughout the clusters. Scale bar, 50 μm.
LpOps5 homologs in each of the extant horseshoe crabs examined, and homologues are present in crustaceans. Therefore, an LpOps5 homologue was likely present in the last common ancestor of crustaceans and euchelicerates, but was lost in the lineage leading to scorpion and spiders.

Long-wavelength sensitive opsin genes have expanded in Limulus by apparent tandem gene duplications to form the LpOps1–4 cluster. This expansion may be of particular functional significance for Limulus vision. LpOps1–4 is co-expressed with LpOps5 in rhabdomes of Limulus LE retinular cells, but the proteins encoded by the LpOps1–4 genes are ~4 times more abundant than those encoded by LpOps5 (Katti et al. 2010). Furthermore, while the LpOps5 protein concentration in rhabdomes is relatively stable day-to-night, probably because of a steady rate of turnover, the LpOps1–4 protein concentration in rhabdomes changes dramatically. It falls to 50% of its nighttime peak early in the day in response to the onset of light, and it is restored to its nighttime peak concentration by 4 h after sunset in response to darkness and signals from an internal circadian clock (Battelle 2013; Battelle et al. 2013). In addition, LpOps1–4-containing rhabdome membranes are actively shed and renewed throughout the day (Sacunas et al. 2002; Katti et al. 2010; Battelle et al. 2013). This indicates that processes involved in LpOps1–4 protein turnover are particularly active and highly regulated. They are thought to contribute to a dramatic nighttime increase in the sensitivity of the LE to light (Barlow et al. 1977), and thus the animal’s ability to find mates while spawning at night (Barlow et al. 1982).

The eyes of some species of scorpions and spiders also show dramatic changes in sensitivity that are controlled in part by a circadian system similar to that in Limulus (Fleissner and Heinrichs 1982; Fleissner 1983; Fleissner and Fleissner 1988; Yamashita 2002). These are correlated with daily changes in rhabdome structure (Fleissner and Fleissner 1988; Grusch et al. 1997), but in scorpions and spiders it is not yet known whether changes in sensitivity and structure correlate with changes in opsin protein concentrations in rhabdome membranes. It is interesting to note, however, that, like Limulus, the nocturnal spider S. mimosaum (Crouch and Lubin 2000) has an expanded repertoire of LWS opsin (fig. 2).

Insights from Gene Structure

The structure of Limulus opsin genes confirms their placement in phylogenetic trees based on sequence homology and provides new insights into the relationships among opsins. Gene structure is classically considered highly conserved in evolution and therefore a feature that can provide insights into relationships among gene families (Rokas and Holland 2000). The present study and that of Dalal et al., (2003) are the first to describe genomic structures for chelicerate opsins.

Our finding that the introns of LpCOps1 and 2 are strictly conserved with introns 1, 3 and 4 of all known vertebrate C-type opsins and with introns in C-type opsin genes of other invertebrates including insects (e.g., Tribolium), and crustaceans (e.g., D. pluex) (Fridmanis et al. 2007), supports their placement within the large C-type opsin group and extends the idea of the homology of all bilaterian C-type opsins. Similarly, our placement of LpPerOps1 among the peropsins within the RGR/Go group is supported by our finding that introns 1, 3 and 5 of the LpPerOps1 gene are identical to those in peropsin genes of vertebrates, hemichordates and several other invertebrates (Albalat 2012). The conserved introns we identified in Limulus SWS opsins and arthropins (fig. 2) and recovered in R-type opsins from diverse species (supplementary fig. S4, Supplementary Material online), are consistent with the placement of these opsins within the R-type opsin group and suggest the structure of Limulus SWS and arthropin genes is ancient among R-type opsins.

Surprisingly, Limulus LWS opsin genes, which are closely related to the SWS opsins based on sequence homology and phylogenetic placement, have a very different structure (fig. 3). However, the structure of Limulus LWS R-type opsin genes matches that of LWS R-type opsin genes in insects (supplementary fig. S6, Supplementary Material online). This finding supports placement of LpOps1–4 and 8 among other LWS R-type opsins and suggests the structure of LpOps1–4 is ancient among arthropod LWS R-type opsins.

Our analyses of gene structure points further to a relationship between LWS LpOps1–4 and LpPerOps1. The introns in LWS LpOps1–4 align with and match the phase of introns 4 and 5 in LpPerOps1. As was described above, the introns of LpOps1–4 are probably deeply rooted in the phylogeny of LWS arthropod opsins, and intron 5 in LpPerOps1 is highly conserved in all peropsins. The apparent relationship between LWS opsins and peropsins revealed by gene structure was unexpected based on the phylogeny shown in fig. 2, although this relationship has been suggested in previous studies (Porter et al. 2012). Together these data show remarkable conservation of several intron-exon boundaries, which both support our phylogenetic classifications, and also suggest higher order relationships between major classes of opsins. In addition, the high level of intron conservation suggests a highly conserved regulatory role of these introns (e.g., mRNA stability, nuclear transport, etc.).

Curiously, LpOps5 was the only chelicerate opsin in a clade originally considered unique to crustaceans (Kashiyma et al. 2009). In the present study, we added to this MWS clade two LpOps5 homologues from other extant horseshoe crab species; thus strengthening the idea that an LpOps5 homolog was present in the last common ancestor of arthropods (Henze and Oakley 2015). This idea is further supported by our observation that the LpOps5 gene has an intron in common with some Daphnia MWS opsins (supplementary fig. S7, Supplementary Material online).
Expression Pattern of the Full Repertoire of Limulus Opsins

Eyes

In previous studies, we showed by RT-PCR and immunocytochemistry or in situ hybridization that ten different opsins are expressed in eyes. The RT-PCR screens in the present study added LpOps9 to the list of opsin transcripts detected in each eye type and LpUVOps2 and LpArthops1 to the list in larval eyes. However, we were unable to verify their expression by in situ hybridization. These findings are reminiscent of results from our previous expression studies of LpOps1–4 and 5 in MEs. In MEs, we routinely detected LpOps1–4 and 5 transcripts by RT-PCR (Smith et al. 1993; Katti et al. 2010; Battelle et al. 2015) but not by in situ hybridization (Katti et al. 2010; Battelle et al. 2015). We were also unable to detect LpOps1–4 or 5 proteins in MEs by immunocytochemistry using antibodies that routinely detect them in photoreceptors of LEs and larval eyes (Katti et al. 2010; Battelle et al. 2015). Thus, the functional significance of opsin transcripts detected by RT-PCR only must be viewed with caution, especially in tissues where other opsins are clearly expressed.

CNS

We anticipated finding LpOps1–4, and 5, LpUVOps1 and LpPerOps1 transcripts in the brain because all are expressed in VE photoreceptor cell bodies, and these are often located on the anterior brain (fig. 6 and Chamberlain and Wyse 1986). Indeed, when we assayed for opsin transcripts in cDNA prepared from brains from which most VE photoreceptor cell bodies had been removed by cutting away a portion of the anterior brain, we did not detect LpUVOps1. This suggests VE photoreceptor cell bodies are the source of LpUVOps1 transcripts in the brain, and as we did not detect LpUVOps1 transcripts elsewhere in the CNS or tail, we conclude that LpUVOps1 is eye-specific. By contrast LpOps1–4, 5 and LpPerOps1 transcripts were detected in cDNA from brains lacking most ventral photoreceptor cell bodies (supplementary fig. S8, Supplementary Material online) suggesting these transcripts are also present elsewhere in the brain.

Our in situ hybridization assays revealed that major sources of LpOps1–4 transcripts in the brain are the axons of lateral and VE photoreceptors (figs. 18, 6A and B). The in situ labeling of LpOps1–4 transcripts in these axons was so intense that it resembled labeling seen in tract-tracing studies of lateral and ventral optic nerves (Chamberlain and Barlow 1980; Calman and Battelle 1991; Calman et al. 1991) revealing photoreceptor axon terminals in optic ganglia. The intense labeling of LpOps1–4 transcripts was particularly surprising in lateral optic nerves because in the older juveniles we used in our in situ hybridization assays, the lateral optic nerve is ~15 mm long. Active transport of transcripts long distances to axon terminals is commonly observed for transcripts encoding proteins that are translated and present in axons and terminals (Giuditta et al. 2008). However, this mechanism does not seem relevant to LpOps1–4 transcripts, because LpOps1–4 proteins are detected only in the specialized rhabdomeric membranes in photoreceptor cell bodies. The presence of LpOps1–4 transcripts in lateral and ventral optic nerves may reflect a particularly high level of these transcripts in photoreceptors, perhaps a consequence of transcription from all four genes in the LpOps1–4 gene cluster. High LpOps1–4 transcript levels are also consistent with the finding mentioned above that LpOps1–4 proteins are ~4 times more abundant in LE photoreceptors than LpOps5 (Katti et al. 2010), which is encoded by a single gene.

Lateral eyes and VE axons may also be the sources of LpOps5, 9, 10, LpUVOps2 and LpArthops1 transcripts in brain as each is present in eye photoreceptors (fig. 5). Our inability to detect them in optic nerves by in situ hybridization suggests their levels in axons are low. ME photoreceptors expressing LpOps6, 7, and 8 also project to the brain (fig. 1B), but these transcripts are not detected in brain even by RT-PCR. If LpOps6, 7 and 8 transcripts are present in median optic nerves where they enter the brain, their levels must be extremely low. This may be because only ~30% of ME photoreceptors express these opsins (Nolte and Brown 1972; Battelle et al. 2015).

Axons from photoreceptors in eyes cannot be the only source of opsin transcripts detected by RT-PCR in the brain and elsewhere in the CNS and tail. For example, we detected C-type opsin transcripts in the brain, and they are not expressed in any of the eyes. Furthermore, the opsins we detected in the synganglion, segmental ganglia and tail cannot be explained by input from eyes. A clear concern is that, except for LpPerOps1, which is discussed separately below, we were unable to identify any opsin expressing cells in the CNS by in situ hybridization. This raises questions about their functional relevance. On the other hand, some of the opsins we detected in the CNS and tail, even though expressed at low levels, may contribute to the extracocular photosensitivity that has been described in Limulus.

Extraocular Photosensitivity

No photosensitive cells have been identified in the brain, except for ventral larval eye photoreceptors located at the brain, and no photosensitive cells have been identified in the synganglion. This may be because there has been no systematic search for photosensitive cells in these tissues. Photosensitive cells have been described in segmental ganglia (Mori et al. 2004), and there is good evidence that the tail is...
photosensitive (Hanna et al. 1988; Renninger et al. 1997). The question most relevant to the current study is: Which of the opsins expressed in these tissues might contribute to this extracocular photosensitivity?

Photosensitive cells in segmental ganglia were identified with intracellular recordings which showed that each ganglion contains one or several photoreceptors, and that all photoreceptors so far examined are maximally sensitive to light at 425 nm (Mori et al. 2004). This suggests that LpOps1–4 and 5, with maximum sensitivities at ~520 nm, and LpUVOps2 are not involved. The remaining candidates are the C-type opsins, LpOps9 and LpArthops1. Photosensitivity in the tail was demonstrated by showing that the phase of the animal’s circadian clock can be shifted by illuminating the tail with broad spectrum light (Hanna et al. 1988). As the spectral sensitivity of the response was not investigated further, all of the opsins we detected in the tail are potentially involved. However, LpOps10 is of particular interest because its transcripts are tail-specific.

**Peropsin**

Peropsin is expressed in eyes and throughout the CNS. The results described here provide the first clear example of peropsin expression outside of eyes. Peropsin proteins were previously identified in the eyes of mammals (Sun et al. 1997) and other vertebrates (e.g., Bailey and Cassone 2004), in some but not all spider eyes (Nagata et al. 2010; Eriksson et al. 2013) and in Limulus eyes (Battelle et al. 2015). Peropsin transcripts have been detected in spider brain (Eriksson et al. 2013) and in transcriptomes of crustaceans, myriapods and insects (Henze and Oakley 2015), but the cellular distributions of these transcripts are not known. Where peropsin distribution has been examined, it is consistently found in glia or pigment cells most often, but not exclusively, surrounding photoreceptors. Based on its frequent association with photoreceptors, peropsins have been postulated to play a role in vision, possibly as a photoisomerase (Sun et al. 1997; Chen et al. 2001; Nagata et al. 2010).

Peropsins may function in vision, but the broad distribution of LpPerOps1 in the CNS we observe suggests it has other functions as well. Because the distribution of LpPerOps1 is not uniform in the CNS, we considered it might be specifically associated with opsin-expressing cells. But we think this unlikely. Nothing is known about the distribution of photosensitivity in the synganglion, but it seems unlikely that all the cells surrounded by LpPerOps1-expressing glia in this ganglion are photoreceptors. The number of cells surrounded by LpPerOps1-expressing glia in segmental ganglia also seems much larger than the 2% of cells thought to be photosensitive (Mori et al. 2004). Our findings add to the puzzle of peropsin function. Clearly, many more studies are required to clarify its role in eyes and in the CNS.

**Conclusions**

These analyses of the full repertoire of opsins from the *Limulus* genome provide unprecedented insight into the visual system of a chelicerate. As such, these data are key to reconstructing the most recent common ancestor of arthropods and providing fundamental evolutionary insights into processes that shaped the immense diversity of visual systems found in Arthropoda.

**Supplementary Material**

Supplementary tables S1–S4 and figs. S1–S8 are available at Genome Biology and Evolution online (http://www.gbe.oxfordjournals.org/).

**Acknowledgments**

The *Limulus* genome project was conceived by N.P. and R.K.W. and completed at the McDonnell Genome Institute at Washington University School of Medicine by W.C.W., P.M., M.J.M. and L.F. with critical input from P.J.G., S.A.S., M.E.P. and B.-A.B. The project was supported by the National Human Genome Research Institute [2U54HG00307909 to R.K.W.], and the National Science Foundation [MCB1021636 to P.J.G.]. Analyses of *Limulus* opsins and their expression was conceived by B.-A.B., J.F.R. and M.L.P. and executed at the University of Florida’s Whitney Laboratory for Marine Bioscience and the University of Hawaii, Manoa by B.-A. B., J.F.R., M.L.R., K.E.K., S.R.S., and C.E.M. The study was supported by the National Science Foundation [IOS 1146175 and DBI 0648969 to B.-A.B.], startup funds from the University of Florida DSP Research Strategic Initiatives #00114464 and University of Florida Office of the Provost Programs to J.F.R., and startup funds from the University of Hawaii, Manoa to M.L.P. S.R.S was a National Science Foundation Research Experience for Undergraduates trainee at the Whitney Laboratory [DBI 0648969]. The article was prepared by B.-A.B., J.F.R., M.L.P. and M.J.M. with critical input from each of the authors. The authors thank Terence Murphey at NCBI for genome annotation.

**Literature Cited**


Associate editor: Belinda Chang