EARLY EVOLUTION OF THE GENETIC BASIS FOR SOMA IN THE VOLVOCACEAE

Erik R. Hanschen,1,2 Patrick J. Ferris,1 and Richard E. Michod1

1Department of Ecology and Evolutionary Biology, University of Arizona, 1041 E. Lowell St., Biosciences West, Rm. 310, Tucson, Arizona
2E-mail: hanschen@email.arizona.edu

Received September 30, 2013
Accepted March 16, 2014

To understand the hierarchy of life in evolutionary terms, we must explain why groups of one kind of individual, say cells, evolve into a new higher level individual, a multicellular organism. A fundamental step in this process is the division of labor into nonreproductive altruistic soma. The regA gene is critical for somatic differentiation in Volvox carteri, a multicellular species of volvocine algae. We report the sequence of regA-like genes and several syntenic markers from divergent species of Volvox. We show that regA evolved early in the volvocines and predict that lineages with and without soma descended from a regA-containing ancestor. We hypothesize an alternate evolutionary history of regA than the prevailing “proto-regA” hypothesis. The variation in presence of soma may be explained by multiple lineages independently evolving soma utilizing regA or alternate genetic pathways. Our prediction that the genetic basis for soma exists in species without somatic cells raises a number of questions, most fundamentally, under what conditions would species with the genetic potential for soma, and hence greater individuality, not evolve these traits. We conclude that the evolution of individuality in the volvocine algae is more complicated and labile than previously appreciated on theoretical grounds.

KEY WORDS: Gene duplication, germ-soma differentiation, major transitions, multicellularity, regA, Volvox.

The familiar hierarchy of life consists of different kinds of evolutionary individuals, for example, prokaryotic cells, eukaryotic cells, multicellular organisms, and eusocial societies. During evolutionary transitions in individuality, groups of individuals become a new kind of individual and the level of selection is transferred from the individual to the group—afterwards a new kind of individual. A fundamental characteristic of these transitions is the functional integration of the former evolutionary individuals into a new indivisible group. During this process, fitness is reorganized so that the former lower level individuals specialize in the fitness components of the group and the unit of fitness changes from the lower level to the new higher level individual (Michod 1999).

Fitness involves two basic components, reproduction and viability. During the evolution of multicellularity, functional integration occurs when cells relinquish their total fitness by specializing in one of the two basic fitness components of the cell group (Michod 2007). In numerous independent evolutions of multicellularity across the tree of life, fitness specialization and integration often takes the form of cellular differentiation into germ and soma cells, where germ cells specialize on reproduction and somatic cells specialize on survival (Bonner 1998; Michod 2007; Simpson 2012). Because these cells specialize on only one fitness component, they can no longer both survive and reproduce on their own, were they to leave the group. As a result, the group becomes indivisible and hence an individual.

Understanding the evolution of reproductive differentiation and specialization during the evolution of multicellularity requires an understanding of the genetic basis underlying these traits as well as their ecological and phylogenetic context. For example, in the volvocine algae studied here, a gene critical for somatic differentiation in Volvox carteri, regA, is homologous to a gene in a unicellular relative, Chlamydomonas reinhardtii. In C. reinhardtii, this homologous gene is expressed in stressful environments in which growth and reproduction should be down regulated.
(Nedelcu and Michod 2006; Nedelcu 2009). Thus, it appears that a life-history gene for down regulating reproduction in stressful environments in a unicellular ancestor could have been co-opted to down regulate reproduction in somatic cells, by changing its expression from a temporal to spatial context (Michod 1999; Michod and Roze 1999; Schlichting 2003; Nedelcu and Michod 2006; Nedelcu 2009).

The volvocine algae are a clade of green algae including unicellular (e.g., C. reinhardtii), colonial (e.g., Gonium pectorale), and multicellular species (e.g., V. carteri). The volvocine algae do not have a unicellular ancestor and unlike other model systems for the evolution of multicellularity and cellular differentiation, have a relatively recent origin of multicellularity which allows identification of the genetic changes associated with the steps to multicellularity (Kirk 2005; Herron et al. 2009; Prochnik et al. 2010; Leliaert et al. 2012). In V. carteri, germ and soma differentiation mutants produce a variety of developmental phenotypes such as the reg or “regenerator” mutant phenotype (Starr 1970; Husky and Griffin 1979; Harper et al. 1987; Kirk et al. 1987, 1999; Adams et al. 1990). In this mutant phenotype, somatic cells initially develop apparently normally, but then redifferentiate into asexual reproductive germ cells which divide to produce offspring. All regenerator mutants map to the regA locus and transformational rescue of this mutant with wild-type regA recovers the wild-type phenotype (Kirk et al. 1999; Nishii and Miller 2010). The evolutionary history of regA is of interest to the evolution of reproductive division of labor and the evolution of individuality in this lineage. Relatively, little is known about when and why regA evolved except that the V. carteri lineage is thought to have evolved sterile somatic cells between 76.6 and 116 million years ago (Herron et al. 2009).

The regA gene is a member of the VARL gene family (Volvocine Algae RegA Like) which encodes a single DNA-binding SAND domain, approximately 80 amino acids long. The VARL domain is characterized by a short N-terminal extension separated from the longer core VARL domain by a variable length linker (Duncan et al. 2007). regA is expressed in smaller somatic cells after embryonic development and is thought to be a master regulator transcription factor of nuclear-encoded chloroplast biogenesis genes (Kirk 1998; Kirk et al. 1999; Meissner et al. 1999; Duncan et al. 2006; Nedelcu et al. 2006). regA is hypothesized to prevent cell division by preventing cells from growing by down regulating chloroplast biosynthesis, thereby creating somatic cells (Meissner et al. 1999).

In V. carteri, the regA gene is part of the regA cluster, a series of tandem duplications involving regA, rlsA, rlsB, and rlsC. The RNA expression of regA, rlsA, rlsB, and rlsC peaks during cellular differentiation, soon after embryonic development, suggesting the entirety of the regA cluster plays a role in somatic differentiation (Harryman 2012). These four genes are closely related to another VARL gene, rlsD. Although rlsD in V. carteri and RLS1 in C. reinhardtii are orthologs, there are no orthologs of the V. carteri tandem duplication genes in C. reinhardtii. A phylogenetic analysis found that the regA cluster genes are more closely related to each other than to rlsD (Duncan et al. 2007). This led Duncan et al. (2007) to hypothesize that regA evolved from an ancient “proto-regA” gene which was subsequently lost in C. reinhardtii (Fig. S1).

Here, we investigate the presence of VARL genes in a diverse group of Volvox species: Volvox africanus, Volvox obversus, V. carteri f. weissmarnia, Volvox gigas, and Volvox ferrisii. As Volvox is polyphyletic, these species are both closely related and divergent to each other and to V. carteri. Volvox gigas is thought to share an origin of somatic cells with V. carteri but V. ferrisii diverged earlier (Fig. 1) and is thought to have an independent evolution of somatic cells (Nozaki et al. 2002; Herron et al. 2009). The common ancestor of V. carteri, V. gigas, and V. ferrisii has also given rise to species of the genus Yamagishiella, Eudorina, and Pleodorina (Nozaki et al. 2002, 2006). Yamagishiella species have 32 cells without somatic differentiation (Rayburn and Starr 1974) whereas Eudorina comprises species with 32 cells without somatic cells (Eudorina elegans) and with facultative somatic cells (Eudorina illinoisensis, Goldstein 1964). Pleodorina species vary between 32 and 128 cells, and have between 12.5 and 50% somatic cells (Iyengar and Desikachary 1981). The clade composed of V. carteri and V. ferrisii includes species without somatic cells, with facultative somatic cells, and with obligate somatic cells.

The genus Volvox is diverse, representing four distinct developmental programs designated D1–D4 (Fig. 1, also see Desnitski 1995). In D1, (e.g., V. gigas) germ cells are initially small, grow substantially to their mature size, and then undergo rapid embryonic divisions without cell growth during embryogenesis. D2 (e.g., V. carteri, V. africanus, V. obversus) is characterized by initially large germ cells followed by rapid divisions and unequal divisions in the germ line. D3 (e.g., Volvox tertius) is characterized by slow embryonic divisions and initially large germ cells. D4 (e.g., V. ferrisii) is characterized by initially small germ cells that undergo slow embryonic divisions with cells growing between divisions (Desnitski 1995). To date, the presence and function of VARL genes has only been demonstrated in C. reinhardtii and V. carteri, so the possible existence and role of regA in other developmental programs was unknown.

We report here that both V. gigas and V. ferrisii contain the tandem duplication of VARL genes including regA and that V. ferrisii has an additional gene in this tandem array. On the basis of the results reported here, these tandem duplications are thought to be ancestral to species with and without soma. The phylogenetic distribution of somatic cells in the volvocine algae (Fig. 1) may be explained by independent tandem duplications, multiple lineages independently evolving somatic cells via regA, or multiple
Species phylogeny of the Volvocine algae adapted regA (UTEX 1890), V. carteri V. rousseletii weismannia regA μ V. carteri and the complete regA weismannia (Table S1). cluster genes may also be present (i.e., we have no 3 Sac V. carteri (Miller and Kirk 1999). Plasmid and cosmid DNA was nagariensis and cluster V ARL genes was regA V. ferrisii group and an approximate date of 200 million Ne-G. pectorale CLUSTER ISOLATION SCHEME BY SPECIES 2014 nagariensis regA cluster genes from weismannia f. Pleodorina the last common ancestor of (2009) inferred an approximate date of 100 million years ago for lines independently evolving somatic cells via different genetic pathways.

**Material and Methods**

**ALGAE CULTURES and CULTURE CONDITIONS**

Strains were grown in standard Volvox medium at 25°C on a 16:8 h light:dark cycle, at ~35 μmol photons/m²/s on a rotary shaker. The following strains were used: V. carteri f. weismannia Nebraska female (UTEX 1874), V. obversus male (UTEX 1865), V. africanus (UTEX 1890), V. gigas (UTEX 1895), V. ferrisii (NIES 2739), and V. rousseletii male (UTEX 1861). All strains were genotyped by sequencing the nuclear rDNA ITS after polymerase chain reaction (PCR) amplification (Coleman et al. 1994).

**DNA PREPARATION**

Genomic DNA was prepared using the protocol described for V. carteri (Miller and Kirk 1999). Plasmid and cosmid DNA was prepared using the Qiaprep Spin Miniprep Kit or the Qiagen Plasmid Midi Kit according to the manufacturer’s instructions.

**PCR CLONING OF GENES**

All PCR reactions (with one exception noted below) were performed using 2X Phusion HF Master Mix (Thermo Scientific, Waltham, MA) with 3% DMSO and 25 ng of template when using genomic DNA. Cloning of regA cluster VARL genes was initiated by performing genomic PCR with degenerate primers. Three forward primers (regF1, regF2, regF3) and one reverse primer (regR) were designed using the protein sequences of the regAcluster genes from V. carteri f. nagariensis (Table S1). The regF1 primer was only used successfully with V. obversus. The regF2 and regF3 primers are based on the same amino acid sequence, but regF3 is more degenerate. Cycling conditions for these experiments were 98°C for 3 min, followed by 35 cycles of 98°C/10 sec, 68°C/20 sec, 72°C/60 sec, then a finishing step of 5 min at 72°C. The PCR product produced typically included 57 coding nucleotides (not contributed by the primers) and a conserved intron. PCR amplifications intended for sequencing were performed in at least two (usually more) separate reactions and then combined before sequencing to reduce the possibility of PCR errors.

Inverse PCR was performed following Sambrook and Russell (2001). Approximately 500 ng of genomic DNA were digested with SacI (for V. africanus) or PstI (for V. obversus and V. carteri f. weismannia) and the restriction enzyme was then heat inactivated. Following ethanol precipitation and resuspension of the DNA at a concentration of 2 ng/μl, the restriction fragments were circularized by ligation with T4 DNA ligase (New England Biolabs, Ipswich, MA). The DNA ligase was heat inactivated, the DNA concentrated by ethanol precipitation, then used as template for PCR with inverse PCR primers. The products of inverse PCR were gel-purified (QIAquick Gel Extraction Kit) and their DNA sequence determined.

**regA CLUSTER ISOLATION SCHEME BY SPECIES**

Volvox carteri f. weismannia: genomic PCR with regF2 and regR primers yielded products of approximately 600 and 1100 nucleotides; the sequence of the latter was homologous to the conserved intron in V. carteri f. nagariensis and V. carteri f. kawasakiensis regA genes (Miller et al. 1999; Duncan et al. 2006). PCR primers were used for inverse PCR of genomic DNA to extend the sequence some 750 base pairs further upstream. We then

**Figure 1.** Species phylogeny of the Volvocine algae adapted from Yamada et al. (2008) and Isaka et al. (2012). Tree is rooted (not shown) using G. pectorale (volvocine alga) as the outgroup. Species with sequences reported here are in bold. Evolution of soma (G) and loss of soma (L) as inferred by Herron et al. (2009) are denoted. Likely evolution of regA (R) is denoted. Herron et al. (2009) inferred an approximate date of 100 million years ago for the last common ancestor of Eudorina elegans UTEX 1205 and the Pleodorina/Volvox group and an approximate date of 200 million years ago for the last common ancestor of V. ferrisii and V. carteri (R). Filled circles represent obligate somatic cell differentiation, open circles represent facultative somatic cell differentiation, and dots represent absence of somatic cell differentiation. In the second column, division programs as defined in Desnitski (1995) are shown. In the third column, species with regA and the complete regA cluster are shown. For species with only evidence of regA, the other regA cluster genes may also be present (i.e., we have no evidence of presence or absence of other regA cluster genes for these species).
used a direct genomic PCR approach. The *V. carteri* *f. weismannia* sequence was used to design forward PCR primers that were paired with reverse primers (some slightly degenerate) designed from coding sequences in exon 6 (exons are numbered as for *V. carteri* *f. kawasakiensis* *regA*, Duncan et al. 2006) of the two known *V. carteri* *regA* genes (Cterm1, Cterm3, and Cterm4, see Table S1). *Volvox carteri* *f. weismannia*-specific reverse PCR primers were paired with forward primers from the 5’ end of exon 5 (regNterm) and the 3’ end of exon 1 (int1). The PCR products from the reactions using these five primers were sequenced (by primer walking when necessary). Finally, the promoter region was amplified using a *V. carteri* *f. weismannia*-specific reverse PCR primer from *regA* exon 2 (rlsA2) with a forward primer rlsA1 (designed from rlsA1) and sequenced. The final *V. carteri* *f. weismannia* assembly (Genbank KF555603) extends from the promoter region of *regA* to within a few nucleotides of the predicted stop codon.

*Volvox obversus*: genomic PCR with regF1 and regR yielded a product of ~2200 nucleotides. PCR primers designed from this sequence were used for inverse PCR of genomic DNA to extend this sequence to include the complete nucleotide sequence encoding the VARL domain (Genbank KF545953).

*Volvox africansus*: genomic PCR with regF2 and regR yielded two products of approximately 800 nucleotides. Both PCR products were sequenced; PCR primers were designed from the one most similar to *regA* (*V. carteri* *f. nagariensis*) and used for inverse PCR of genomic DNA to extend this sequence to include the complete nucleotide sequence encoding the VARL domain (Genbank KF555602).

*Volvox gigas*: genomic PCR with regF2 and regR (66°C annealing temperature) yielded multiple products. These were cloned using the pGEM-T Easy Vector System (Promega, Madison, WI) after A-tailing and transformation into NEB 5α competent *Escherichia coli* (New England Biolabs). Plasmid DNA was prepared from ampicillin-resistant white colonies selected on Luria Broth (LB)/ampicillin/IPTG/X-GAL plates, then sequenced using primers based on the T7 and SP6 promoter sequences within the pGEM-T Easy Vector. Sequences encoding part of three different VARL domains were identified among the subclones: rlsA, rlsB, and rlsC. Three oligomer hybridization primers (Table S1) were designed for screening the *V. gigas* cosmid library (see below). The *regA* gene was discovered during cosmid sequencing. *rlsB* appeared to be incomplete at the C-terminus, so the sequence was extended to a stop codon by TAIL-PCR, following Dent et al. (2005) using Taq 2X Master Mix (New England Biolabs). The *V. gigas* sequences are Genbank KF582790 and KF582789.

*Volvox ferrissii*: genomic PCR with regF2/regF3 and regR was performed using an eight-step gradient in annealing temperature (60–72°C) yielding multiple products. The PCR products were pooled, subcloned, and sequenced. The sequence from two of eight subclones from the regF2/regR reactions encoded part of the *regA* VARL domain. The sequence from one of 13 subclones from the regF3/regR reactions encoded part of the *rlsB* VARL domain. An oligomer hybridization primer (Table S1) was designed from each of these for screening the *V. ferrissii* cosmid library (see below). The *regA* probe yielded three overlapping cosmids, one of which also hybridized to the *rlsB* probe. PCR using regF2 or regF3 and regR using cosmid DNA templates identified two products encoding additional VARL domains (*rlsA* and *rlsC*). The *rlsN* and *ackB* genes were identified during cosmid sequencing. An oligonucleotide hybridization primer (Table S1) was designed from the DNA sequence around the starting methionine of *ackB* and used to screen the *V. ferrissii* cosmid library to identify an overlapping cosmid that proved to have the *rlsD* gene as determined by partial sequencing of the cosmid. The 80,171 nucleotide *V. ferrissii* sequence is Genbank KF607039.

*Volvox rousseletii*: two sets of PCR primers (fer6F/fer6R and fer6F2/fer6R2, Table S1) were designed from the *V. ferrissii* *rlsN* sequence and used in PCR to yield overlapping products that span the sequence encoding the two *rlsN* VARL domains in *V. rousseletii* (Genbank KF615907).

**LIBRARY CONSTRUCTION**

Cosmid libraries from *V. gigas* and *V. ferrissii* were constructed from size-fractionated genomic DNA using the pWEB-TNC Cosmid Cloning Kit (Epiconcentre, Madison, WI), following the manufacturer’s instructions. To create amplified pools of the libraries, the *E. coli* host strain EPI100-T1 was transduced with an aliquot of the packaged cosmid library and plated at a density of several thousand colonies per plate. The colonies on individual plates were resuspended separately in LB, DMSO added to 8% (v/v), then stored at ~80°C. An aliquot from each pool was used to inoculate a 5 ml overnight culture in LB/ampicillin (100 μg/ml) from which cosmid DNA was prepared for use as PCR template for ascertaining whether genes of interest were present within a given pool. Pools of interest were plated on LB/ampicillin agar plates at a density of 500–1000 colonies; after overnight incubation, *E. coli* colonies were transferred to nitrocellulose (NitoPure supported nitrocellulose, 0.45 μm, 82 mm circle; Maine Manufacturing, LLC, Sanford, ME), lysed in situ, and the DNA bound to the filter by UV crosslinking (Sambrook and Russell 2001). The nitrocellulose lifts were incubated overnight with radiolabeled probe in a rotary hybridization oven at a temperature 15°C below the Tm of the hybridization primer. The hybridization buffer was 5X SCP (5X SCP is 0.5 M NaCl, 0.05 M EDTA, 0.15 M sodium phosphate, pH 6.6), 1% sodium N-lauroylsarcosine, 100 μg/ml fish sperm DNA, MB-grade (Roche, Basel, Switzerland). Lifts were rinsed at room temperature in 2X SCP, 1% SDS (three
changes, 20 min each), then in 0.5X SCP, 1% SDS (two changes, 10 min each), then autoradiographed. Colonies showing positive hybridization were subcloned and cosmid DNA prepared. Only one cosmid was used from any given pool.

Hybridization primers were end-labeled at 37°C for 60 min in a 20 µl reaction containing 25 pmole of primer, 50 pmole of γ-32P ATP (6000 Ci/mmol; PerkinElmer NEG035C005MC), and 10U polynucleotide kinase (Fermentas, Waltham, MA) in 1X PNK buffer A. Unincorporated ATP was removed using a Micro G-25 Spin Column (Santa Cruz Biotechnology, Dallas, TX).

DNA SEQUENCING

DNA sequencing of PCR products, plasmids, and cosmids was performed by the University of Arizona Genetics Core service using Applied Biosystems 3730 DNA Analyzers (Waltham, MA). DNA sequences were assembled into contigs using the CLC Main Workbench software.

From this assembly, BLAST (version 2.2.26) was used to identify the sequence encoding putative VARL domains which were confirmed with Pfam (version 26.0) and SMART (version 7.0; Altschul et al. 1990; Finn et al. 2010; Letunic et al. 2012). Gene models were obtained with AUGUSTUS (version 2.6.1, trained on C. reinhardtii and using known VARL domains from C. reinhardtii and V. carteri, as hints, Stanke et al. 2006) and manual model building. N-terminal extensions were identified using custom Hidden Markov Models (HMM, HMMER version 3.0, Eddy 1998). Syntenic non-VARL genes were identified by BLAST searches and AUGUSTUS gene modeling.

PHYLOGENETIC ANALYSES

All protein sequences (approximately 80 amino acids) were aligned using MAFFT (version 6.859b) with the accurate L-INS-i option (Katoh et al. 2005). The corrected Akaike information criterion in ProTest (version 3.2, Posada et al. 2011) determined LG+G as the optimal substitution model. Phylogenies were performed using maximum likelihood (ML), maximum parsimony, neighbor joining, and Bayesian methods. ML estimation was implemented using RAxML (version 7.0.4) with a rapid bootstrap analysis, generating 1000 ML replicate trees to estimate bootstrap support (Stamatakis 2006; Stamatakis et al. 2008). Maximum parsimony and neighbor joining were implemented using PHYLIP (version 3.695) with 100 replicate datasets (Felsenstein 1989). For Bayesian analysis, MrBayes (version 3.2.1) was used to run four MCMC replicates for 1 million generations each (Ronquist et al. 2012). The WAG+I+G model was used for direct comparison to the Bayesian tree in Duncan et al. (2007). After a 10% burn-in, trees were sampled every 100 generations to build a 50% majority rule consensus tree. We used codeml in PAML (version 4.3, Yang 2007) to test for stabilizing selection on the sequence encoding the N-terminal extension and VARL domains. A codon-based Z-test for selection using the Nei–Gojobori method with 500 bootstrap replicates was performed on codon-based nucleotide alignments.

Results

PRESENCE OF VARL GENES

We sequenced almost the entire regA gene from V. carteri f. weismannia, as well as a portion of the regA gene of V. obversus and V. africanus that contains the VARL domain. In V. gigas, we sequenced the regA cluster. In V. ferrisii, we sequenced a total of six genes encoding VARL domains. Five of these genes encode a single predicted VARL domain and the last gene encodes two predicted VARL domains. We used Pfam and SMART to predict the presence of a SAND domain (Table S2). We have also sequenced five adjacent genes in V. ferrisii that are syntenic with the regA cluster in both V. ferrisii and V. carteri.

NAMING CONVENTIONS FOR VARL GENES

A standardized nomenclature was used for sequenced genes, following V. carteri f. nagariensis. When the phylogenetic support of a predicted VARL domain clade was not above 0.70 (ML estimation), we used the syntenic relationships compared to V. carteri f. nagariensis to provide names (Table S3). Otherwise, phylogenetic relationships were used to provide gene names. The fourth gene in the regA cluster in V. ferrisii with two predicted SAND domains was named rlsN (with domains rlsNI and rlsN2). Lastly, the syntenic genes found in V. ferrisii were named using conventions from V. carteri f. nagariensis.

STRUCTURE OF THE VARL GENES

We used AUGUSTUS and manual model building to predict exon and intron locations in the sequenced VARL genes in all studied species. In all sequences encoding VARL domains, a single conserved intron position was found. This is position 4 from Duncan et al.’s (2007) analysis. The conserved intron position is further evidence that we have sequenced orthologous genes in V. gigas and V. ferrisii.

The three V. carteri formae (nagariensis, kawasakiensis, and weismannia) are roughly equally diverged and show conservation outside the VARL domain sequence but this conservation is not present in V. africanus and V. obversus. We searched for conservation outside the sequence encoding VARL domains in genes isolated from V. gigas and V. ferrisii but were not able to identify functionally defined, conserved motifs. There are several homologous sequences outside the region encoding the predicted VARL domain in regA, rlsa, rlsB, and rlsC in V. carteri f. nagariensis and V. gigas (Supporting Information Alignments S1–S4). These
short regions provided evidence of orthology and may play a role in facilitating protein–protein interactions but are not currently functionally annotated. This lack of substantial homology outside the VARL domain hinders phylogenetic analysis and is consistent with Duncan et al.’s (2007) attempt to identify conserved domains outside of VARL domains, suggesting that VARL genes indeed encode single-domain proteins.

**PHYLOGENETIC RELATIONSHIPS OF THE VARL GENES**

We made phylogenetic trees using the boundaries of the VARL domain and the N-terminal extension of the VARL domain as defined in Duncan et al. (2007) both for proper comparison to their work and because sequenced genes are consistent with these domain boundaries. The VARL genes we have sequenced are closely related to the functional regA gene in *V. carteri*; however, due to the short sequence of the VARL domain (87 amino acids) there is little phylogenetic support (Fig. 2). To compare our results to those of Duncan et al. (2007), we constructed gene phylogenies using four methods (Bayesian, ML, maximum parsimony, and minimum evolution). Table S6 and S7). We found strong support for the genes we report in *V. obversus*, *V. africanus*, and *V. carteri f. weissmannia* as regA in all four phylogenetic methods, which is consistent with the conserved intron locations as well as short conserved motifs outside the VARL domain (Figs. 2, S2–S4). Our gene phylogenies also found strong support for *rlsA* and *rlsC* in *V. gigas* (Figs. 2, S2–S4).

There are a few notable differences between our analysis and the previous study (Duncan et al. 2007). First, although previous work identified RLS2 in *C. reinhardtii* and *rls* in *V. carteri f. nagariensis* as likely orthologs, there was not strong support for this relationship. In our phylogenetic analyses, RLS2 and *rls* form a well-supported clade, which is consistent with their syntenic relationship. Second, previous work identified RLS8 in *C. reinhardtii* and *rls* in *V. carteri f. nagariensis* as potential orthologs, which is now supported in our analysis. Lastly, Duncan et al. (2007) found strong support for a clade composed of RLS1 in *C. reinhardtii* and *rlsD* in *V. carteri f. nagariensis*, which was the sister group to the regA cluster in *V. carteri f. nagariensis*. In our analysis, this clade of RLS1 and *rlsD* is supported, but its position as the sister group to the regA cluster is not well supported using any phylogenetic method (Figs. 2, S2–S4).

**SYNTENIC RELATIONSHIPS OF THE VARL GENES**

We identified the syntenic structure of the genes in *V. gigas* and *V. ferrisii* (Fig. 3). In *V. carteri f. nagariensis*, regA is part of a tandem duplication including three other closely related VARL genes, *rlsA*, *rlsB*, and *rlsC*. In *V. gigas*, we find that *rlsA*, regA, and *rlsB* are also closely syntenic (Fig. 3). *rlsC* in *V. gigas* is located on a separate, nonoverlapping cosmid, so the linkage relationship of *rlsC* to the rest of the tandem duplication is undetermined.

The syntenic relationships in *V. ferrisii* are more informative (Fig. 3). Four syntenic gene markers (*ido, rpt4, 105104*, and 91985) lie upstream of the regA cluster in *V. ferrisii* and one gene marker, ackB, lies downstream. In *V. carteri f. nagariensis*, ackB is immediately adjacent to *rlsD* but these two genes are located at least 1 Mb away from the regA cluster. In *C. reinhardtii*, ACK2 (the ortholog of ackB) is immediately adjacent to RLS1. In *V. ferrisii*, ackB and *rlsD* are closely linked to the regA cluster.

**STABILIZING SELECTION ON THE VARL GENES**

We used PAML to test for stabilizing selection on the sequence of predicted VARL domains. The gene *rpt4* in *C. reinhardtii*, *V. carteri f. nagariensis*, and *V. ferrisii* (1197 nucleotides) has dN/dS values ranging from 0.030 to 0.050 (Tables S4 and S5). This indicates strong stabilizing selection, which is expected given that *rpt4* is a subunit of the 26S proteasome regulatory polypeptide associated with ubiquitin degradation (Finley 2009). The dN/dS values for the regA genes we have sequenced in related and divergent Volvox species range from 0.00 to 0.378 (261 nucleotides, Tables S6 and S7). dN/dS values for RLS1 and *rlsD* for *C. reinhardtii*, *V. carteri f. nagariensis*, and *V. ferrisii* range from 0.087 to 0.250 (261 nucleotides, Tables S8 and S9). These results indicate that the predicted VARL domains of regA and RLS1 and *rlsD* have been subjected to stabilizing selection, which is consistent with findings in Duncan et al. (2007). This stabilizing selection suggests both regA and *rlsD* in Volvox and RLS1 in *C. reinhardtii* have been functionally conserved. This is consistent with the temporal expression of RLS1 in stressful environments such as light and sulfur deprivation (Nedelcu 2009).

**ANALYSIS OF rlsN**

Our analysis reconstructed *rlsN2* in *V. ferrisii* and *rlsM* in *V. carteri f. nagariensis* as a significant clade. However, because *rlsN* is within the tandem duplication in *V. ferrisii*, *rlsN* is likely the result of an expanded duplication not present in *V. carteri*. Alternatively, five VARL genes may be ancestral and *V. carteri* may have lost *rlsN*. *rlsN* also has a novel gene architecture, containing the sequence of two VARL domains that is not present in *rlsM*. Therefore, we suspect this node does not represent evolutionary relationships, but rather phenomena, such as long branch attraction, that are beyond the scope of this article as this issue does not affect the remainder of our analysis. *rlsN2* is the only domain that SMART did not significantly annotate, so we performed the phylogenetic analyses with and without *rlsN2* and found no difference in phylogenetic support. We did not find a significant N-terminal extension for *rlsN2* with our HMM search. We performed the phylogenetic analyses with and without the N-terminal extension and there was no significant difference.

To investigate potential function of *rlsN* in *V. ferrisii*, we sequenced the region between predicted VARL domains in a closely
 related species, V. rousseletii (Isaka et al. 2012). This region has a $dN/dS$ value of 0.037 between V. ferrisii and V. rousseletii (1176 nucleotides, 96.3% identity). This suggests that the interdomain region is under stabilizing selection and is functionally expressed, although the function of rlsN is unknown. rlsN has a unique gene structure of two predicted VARL domains not previously observed in the volvocine algae. This novel structure suggests unique function which would be worthwhile investigating.

**Discussion**

**VARL GENE ORTHOLOGY**

Previous analyses of the evolutionary history of regA have been restricted to C. reinhardtii and V. carteri, both of which have published genomes (Merchant et al. 2007; Prochnik et al. 2010). Duncan et al. (2007) used these genomes to study the evolutionary history of regA and the VARL gene family and based on their tree topology hypothesized the ancient existence of a "proto-regA" gene which ultimately gave rise to regA in V. carteri and was lost in C. reinhardtii. Because we are unable to replicate strong support for their tree topology, in the next section we consider an alternate evolutionary history of regA.

We have investigated the evolutionary history of regA by sequencing VARL genes from four species of Volvox, both closely related and highly diverged. The sequences from V. africanus, V. obversus, and V. carteri f. weismannia are closely related to previously reported regA sequences in V. carteri f. nagariensis and V. carteri f. kawasakiensis. As all these species have a D2 developmental program, we cannot infer the role of regA in other developmental programs from these species (Fig. 1). In V. gigas (D1) and V. ferrisii (D4), we have demonstrated the presence of the regA cluster. This is the first demonstration of regA genes in a Volvox species with a non-D2 developmental program and equal embryonic cleavage. Under the assumption that these genes are functional, as our $dN/dS$ results suggest, this result demonstrates that regA is not restricted to species with unequal embryonic cleavage, and likely plays a role in somatic differentiation in species with and without unequal embryonic cleavage.
We observe phylogenetic and syntenic relationships between *C. reinhardtii*, *V. carteri*, and *V. ferrisii* (Figs. 2 and 3). The *rlsD* gene and the *regA* cluster in *V. ferrisii* are separated by approximately 7.6 kb including one syntenic marker, *ackB*. This result is in contrast to *V. carteri f. nagariensis* synteny where *rlsD* and the *regA* cluster are separated by at least 1 Mb of sequence (i.e., the minimum distance if scaffolds 23 and 31 in *V. carteri* version 1 are adjacent).

**EVOLUTIONARY ORIGIN OF REGA**

We have discovered that the *regA* cluster (*rlsA, regA, rlsB, rlsN, rlsC*) and *rlsD* are closely linked in *V. ferrisii* and have sequenced five genes that are syntenic with both the *V. carteri f. nagariensis regA* clade and *C. reinhardtii RLS1* (Fig. 3). The *regA* cluster is likely the result of tandem duplications, perhaps by repeated unequal crossing over or backward strand slippage (Chen et al. 2005; Kondrashov and Kondrashov 2006). We hypothesize that the *regA* cluster duplicated from *rlsD* more recently than the speciation of the *Chlamydomonas* and *Volvox* lineages (Fig. 4). In contrast, Duncan et al. (2007) hypothesized an early duplication event giving rise to *regA*, or the entire *regA* cluster, that occurred before the speciation of the *Chlamydomonas* and *Volvox* lineages (Fig. S1). Our data do not allow us to conclusively choose between these alternatives.

Duncan et al. (2007) used an inferred close phylogenetic relationship between *RLS1* in *C. reinhardtii* and *rlsD* in *V. carteri* to hypothesize an ancient ancestor with two VARL genes, the ancestor to *RLS1* and *rlsD*, which we term “proto-RLS1” and proto-*regA* (Fig. S1). Under this hypothesis, the ancestor proto-RLS1 became the orthologous genes *RLS1* and *rlsD* after the *Chlamydomonas* and *Volvox* lineages speciated. Duncan et al. (2007) hypothesized that the proto-*regA* gene was lost in the *C. reinhardtii* lineage but not the *Volvox* lineage thereby producing their observed gene phylogeny.

However, with increased sampling of the *regA* cluster in *V. gigas* and *V. ferrisii*, we have been unable to replicate strong support for the close phylogenetic relationship of *RLS1* and *rlsD*, or for the clade containing *RLS1* and *rlsD* as the sister group to the *regA* cluster (Figs. 2, S2–S4). The existence of a clade containing *RLS1* and *rlsD* sister to the *regA* cluster observed by Duncan et al.
EVOLUTION OF THE GENETIC BASIS FOR SOMA

Figure 4. Model schematic for the revised evolutionary origin of regA. VARL genes are indicated by gray boxes, marker genes are indicated by black boxes, and larger synteny flanking regions as black rectangles. Not to scale. Evolutionary events such as speciation, gene duplication, and recombination are denoted in hypothesized order, the order of events 4a and 4b is unknown.

The proto-regA hypothesis posits an early evolution of regA and predicts that regA should be found in a variety of algae between Volvox and Chlamydomonas. On the other hand, we hypothesize a later evolution of regA from rlsD and predict that rlsD, but not regA, will be found in other unicellular and colonial volvocine algae. The hypotheses would be conclusively tested by sequencing the genomes of colonial volvocine algae such as Tetrabaena socialis or G. pectorale as these are early-diverging colonial species.

IMPLICATIONS FOR MULTICELLULARITY

There are many species with and without somatic cells in the clade containing V. ferrisii and V. carteri (Nozaki et al. 2002, 2006). These include species of Yamagishiella, Eudorina, and Pleodorina, which contain species with no somatic cells, facultative somatic cells, and obligate somatic cells (Fig. 1). Previous

(2007), which we have been unable to confirm with increased sampling of Volvox species, was critical to the development of their “proto-regA” hypothesis.

Due to the low support of this aspect of our phylogenetic trees, our hypothesis of more recent duplication events is not explicitly based on gene phylogenies as was the proto-regA hypothesis. We also note that our observed phylogenetic trees (Figs. 2, S2–S4) do not reconstruct the topology of our revised evolutionary history either, which predicts RLS1 as the outgroup to the regA cluster and rlsD. We do not think this is a problem because stabilizing selection on RLS1 and rlsD (Table S8) and the limited conserved nucleotide sequence (261 nucleotides) available for phylogenetic analysis may cause the aforementioned uncertainty of this aspect of the tree topology (Figs. 2, S2–S4).

Understanding when the genetic basis for somatic cells evolved will shed light into how and why co-option of a life-history gene for suppression of reproduction may have occurred during the evolution of multicellularity. Our revised evolutionary history is consistent with the hypothesis that regA evolved by co-option of a life-history gene in a unicellular ancestor by changing its expression from a temporal to spatial context (Nedelcu and Michod 2006; Nedelcu 2009). Although Duncan et al. (2007) hypothesize regA, with unknown function, evolved in a unicellular species, under our hypothesis, regA evolved in a colonial alga where spatial differentiation may have been beneficial (Herron et al. 2009).

The proto-regA hypothesis posits an early evolution of regA and predicts that regA should be found in a variety of algae between Volvox and Chlamydomonas. On the other hand, we hypothesize a later evolution of regA from rlsD and predict that rlsD, but not regA, will be found in other unicellular and colonial volvocine algae. The hypotheses would be conclusively tested by sequencing the genomes of colonial volvocine algae such as Tetrabaena socialis or G. pectorale as these are early-diverging colonial species.

IMPLICATIONS FOR MULTICELLULARITY

There are many species with and without somatic cells in the clade containing V. ferrisii and V. carteri (Nozaki et al. 2002, 2006). These include species of Yamagishiella, Eudorina, and Pleodorina, which contain species with no somatic cells, facultative somatic cells, and obligate somatic cells (Fig. 1). Previous
phylogenetic analyses have predicted that somatic cells are ancestral to *Eudorina* and *Pleodorina* but not *Yamagishiella* (Herron et al. 2009). However, based on our results, a genetic component thought to be essential to somatic differentiation, the *regA* cluster, is ancestral to all of these species (Fig. 1).

Although the genetic basis for soma is ancestral to lineages with and without somatic cells, it does not mean the common ancestor had somatic cells. Herron and Michod (2008) used ancestral state reconstructions to infer that the last common ancestor of *V. ferrisii* and *V. carteri* did not have somatic cells and predicted that two lineages, the Euvolovix (*V. ferrisii*) and the last common ancestor of *V. gigas* and *V. carteri*, independently evolved somatic cells. This prediction, supported by hypothesis testing, was reached when cellular differentiation was treated as an unordered three-state character (undifferentiated, differentiated soma cells, differentiated soma and germ cells), but not when somatic cells were modeled as a separate character from differentiated germ cells. Consequently, the possibility that somatic cells are ancestral to all species included in this analysis must be acknowledged. If somatic cells are ancestral to all these species, it remains to be seen if, and how, the *regA* cluster is functional in species which have lost soma. The *regA* cluster may be facilitating a spatial cue to undifferentiated cells or a temporal cue based on environmental change (Nedelcu and Michod 2006; Nedelcu 2009; Arakaki et al. 2013).

Either way, our results imply a phylogenetic discordance between somatic phenotype and genotype. If somatic cells are not ancestral to the species studied here, the observation that the somatic genotype, *regA*, is ancestral to species without somatic cells raises three interesting possibilities regarding the regulation of soma: (1) both the *V. carteri* and *V. ferrisii* lineages regulate somatic cells via *regA*, (2) the *V. ferrisii* lineage does not regulate somatic cells via *regA*, or (3) as in (1) the *regA* cluster in *V. carteri* and *V. ferrisii* is used for soma but these clusters represent independent, recent tandem duplications of *rlsD*.

Under (1), the *V. carteri* and *V. ferrisii* lineages have evolved the use of *regA* for somatic regulation in parallel from an ancestor containing *regA* but not soma. This alternative would highlight the role of repeatability and convergence in the evolution of soma. Reproductive altruism is thought to be a fundamental property of all multicellular individuals and eusocial groups. Although recent experimental work has investigated how repeatability impacts evolution (Blount et al. 2008; Meyer et al. 2012), this possibility would suggest a more important role of convergence in the evolution of individuality than previously appreciated.

Under (2), as in (1) there is parallel phenotypic evolution of soma, but *V. ferrisii* uses some other genetic means for somatic regulation. This would also be an important result as it suggests multiple genetic pathways for evolving somatic cells in the volvocine algae. Such a result would suggest that evolving somatic cells is easier than previously understood, as multiple genetic pathways are available for complex traits such as cellular differentiation and individuality (Ardent and Reznick 2008).

For alternative (3), it is known that independent tandem duplications have important effects on evolutionary processes such as host–parasite interactions and life cycle evolution (Hughes and Friedman 2003; Danchin et al. 2010; Cooley et al. 2011), so the possibility of an independent tandem duplication of *rlsD* producing the *regA* cluster in *V. ferrisii* cannot be ignored. However, given the syntenic position of the *regA* cluster and similar gene orientation in *V. carteri* and *V. ferrisii* (Fig. 3), we do not think this likely. If independent duplications did occur, it would have important implications on the likelihood of evolving somatic differentiation due to the constrained number of genetic pathways and specific mutations required for this pathway.

Although we are not currently able to differentiate among these explanations, under all of these possibilities, there have been multiple events regarding the history of somatic cells and individuality in the volvocine algae. Our analysis demonstrates the availability of *regA* as the genetic basis for soma in both *V. carteri* and *V. ferrisii* lineages as well as lineages without somatic cells. Functional analysis of *regA* in *V. ferrisii* could differentiate between explanations (1) and (2). Detailed analysis of VARL gene presence or absence, and subsequent function, in colonial species such as *Yamagishiella unicoCCA* and *E. elegans* is important to testing possibility (3). If these species all have the *regA* cluster (even if pseudogenized), hypothesis (3) would be disproven. We are currently investigating these possibilities.

The first two possibilities are particularly interesting given the number of lineages that did not evolve somatic cells despite their ancestor containing *regA* (R, Fig. 1) and thereby have not further increased in complexity and individuality. Why have these species not evolved soma given the genetic basis for soma? Were the costs of soma simply too great, the required mutations not present, or the immediate advantages nonexistent in their environments? Identifying the environment in which somatic cells are not advantageous would be valuable to understanding the evolution of multicellularity and individuality more generally. If somatic cells are ancestral to the species studied here, what function, if any, has the genetic potential for soma taken on in these species? Answers to these questions will be needed to fully understand the evolution of multicellularity and individuality in this group. Nevertheless, our results imply that individuality, the trait which underlies the levels in the hierarchy of life, is more open to evolutionary change than previous theoretical work envisioned (Buss 1987; Maynard Smith and Szathmáry 1995; Michod 1999).
ACKNOWLEDGMENTS

We would like to thank Q. Li, C. Dieckmann, T. Mittelmeier, M. Barker, and B. J. S. C. Olson for technical help, D. Shelton, Z. Grochau-Weirtz, and S. Miller for discussions and comments, J. Umen for genomic DNA, H. Nozaki for providing algae strains, and three reviewers for helpful comments. This work was supported by the National Science Foundation (grant numbers DEB-0742383, DGE-0654435); and the National Aeronautics and Space Administration (grant number NNX13AH41G).

LITERATURE CITED


Associate Editor: L. Fishman

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher’s website:

Table S1. Sequences of relevant oligonucleotides.

Table S2. E-values of sequenced VARL domains for both Pfam and SMART gene annotation methods.

Table S3. Naming conventions used for specific gene names.

Table S4. dN/dS values for rps4.

Table S5. dN (lower triangle) and dS (upper triangle) values for rps4.

Table S6. dN/dS values for regA.

Table S7. dN (lower triangle) and dS (upper triangle) values for regA.

Table S8. dN/dS values for RLS1 and rlsD.

Table S9. dN (lower triangle) and dS (upper triangle) values for RLS1 and rlsD.

Figure S1. Model schematic for the previous hypothesis for the evolutionary origin of regA (Duncan et al. 2007).

Figure S2. Phylogenetic relationships of VARL domains using Bayesian methods, using identical protein sequences as Figure 1.

Figure S3. Phylogenetic relationships of VARL domains using maximum parsimony methods, using identical protein sequences as Figure 1.

Figure S4. Phylogenetic relationships of VARL domains using neighbor joining methods, using identical protein sequences as Figure 1.