First posted online on 3 January 2017 as 10.1242/dev.141523
Access the most recent version at http://dev.biologists.org/lookup/doi/10.1242/dev.141523

The *Ectocarpus IMMEDIATE UPRIGHT* gene encodes a member of a novel family of cysteine-rich proteins that have an unusual distribution across the eukaryotes

Nicolas Macaisne[1], Fuli Liu[1,†], Delphine Scornet[1], Akira F. Peters[2], Agnieszka Lipinska[1], Marie-Mathilde Perrineau[1], Antoine Henry[1], Martina Strittmatter[1], Susana M. Coelho[1], J. Mark Cock[1,*]

[1]CNRS, Sorbonne Université, UPMC University Paris 06, Algal Genetics Group, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074, F-29688, Roscoff, France, [2]Bezhin Rosko, 29250, Santec, France

[*]Author for correspondance (cock@sb-roscoff.fr)

[†]Current address: Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Qingdao, 266071, China

KEY WORDS: Brown algae; *Ectocarpus*; horizontal gene transfer; initial cell division; life cycle; virus

ABSTRACT

The sporophyte generation of the brown alga *Ectocarpus* sp. exhibits an unusual pattern of development compared to the majority of brown algae. The first cell division is symmetrical and the apical/basal axis is established late in development. In the *IMMEDIATE UPRIGHT* mutant the initial cell undergoes an asymmetric division to immediately establish the apical/basal axis. We provide evidence which suggests that this phenotype corresponds to the ancestral state of the sporophyte. The *IMM* gene encodes a protein of unknown function, containing a repeated motif also found in the *EsV-1-7* gene of the *Ectocarpus* virus. Brown algae possess large families of EsV-1-7 domain genes but these genes are rare in other stramenopiles suggesting that the expansion of this family may have been linked with the emergence of multicellular complexity. EsV-1-7 domain genes have a patchy distribution across eukaryotic supergroups and occur in several viral genomes, suggesting possible horizontal transfer during eukaryote evolution.

INTRODUCTION

Multicellular organisms with haploid-diploid life cycles are found in several major eukaryotic groups including the green lineage (Archaeplastida) and the red and brown macroalgae (Rhodophyta and Phaeophyceae, respectively). In these organisms, a single genome provides the genetic information to deploy two different developmental programs during the course of the life cycle, leading to the construction of the sporophyte and gametophyte generations, respectively (Cock et al., 2013; Coelho et al., 2007). One consequence of this type of life cycle is that the emergence of developmental innovations for one generation of the life cycle can occur without it being necessary to evolve novel developmental regulatory modules *de novo*. This is because it is also possible to adapt regulatory modules that have evolved to function during one generation of the life cycle to carry out related functions during the other generation. An important objective of the developmental biologists that study these organisms has been to understand the relative contributions of these two processes - developmental innovation and trans-generation co-option - to the evolution of multicellularity in these species (Dolan, 2009; Pires and Dolan, 2012; Shaw et al., 2011). In the green lineage, embryophytes (which have haploid-diploid life cycles) are thought to have evolved from a green algal ancestor with a haploid life cycle by the addition of a sporophyte generation (Bower, 1890; Celakovsky, 1874; Dolan, 2009; Haig and Wilczek, 2006; Niklas and Kutschera, 2010; Qiu, 2008; Qiu et al., 2006). It has been proposed that the regulatory networks that controlled the development of early embryophyte sporophytes were recruited to a large extent from the gametophyte generation (Dolan, 2009; Niklas and Kutschera, 2010). Support for this viewpoint has come both from broad comparisons of gametophyte and sporophyte transcriptomes (Nishiyama et al., 2003; Szovenyi et al., 2011) and from demonstrations that homologues of key regulatory genes in embryophyte sporophytes play important roles in gametophyte function in bryophytes (Aoyama et al., 2012; Kubota et al., 2014; Menand et al., 2007; Nishiyama et al., 2003). There are however exceptions (Szovenyi et al., 2011). For example, the KNOX family of TALE homeodomain transcription factors are not expressed during the gametophyte generation in bryophytes and therefore appear to have evolved as sporophyte developmental regulators (Sano et al., 2005).

To more fully understand the relative contributions of developmental innovation and trans-generation co-option to the evolution of multicellularity in organisms with haploid-diploid life cycles it would be of interest to investigate this phenomenon in several lineages that have independently evolved complex multicellularity. Not only would this allow the generality of inferences from studies of the green lineage to be accessed but it would also

allow an evaluation of the importance of the ancestral state on subsequent evolutionary events. For example, the brown algae (Phaeophyceae) most probably evolved from an ancestor that alternated between simple, filamentous sporophyte and gametophyte generations (Kawai et al., 2003; Silberfeld et al., 2010). If this was the case, then the evolution of novel regulatory systems may have played a more important role in the emergence of novel developmental mechanisms than co-option of regulators across generations in this phylogenetic group. Unfortunately, very little is currently known about developmental processes in the brown algae and, for example, no developmental regulatory genes have so far been characterised at the molecular level. However, the recent emergence of the filamentous alga *Ectocarpus* sp. as a model organism for this group (Cock et al., 2013; Cock et al., 2015; Coelho et al., 2012a) has created a context in which this type of question can be addressed.

Ectocarpus sp. has a haploid-diploid life cycle that involves alternation between two generations, which both consist of uniserate filaments with a small number of different cell types and bearing simple reproductive structures (Cock et al., 2015). The morphological similarity of the two generations has allowed mutants affected both in switching between generations (Coelho et al., 2011) and in generation-related developmental processes (Peters et al., 2008) to be isolated. The *immediate upright* (*imm*) mutant is particularly interesting because it has major effects on the early development of the sporophyte generation but causes no visible phenotype during the gametophyte generation (Peters et al., 2008). In individuals that carry this mutation, the initial cell of the sporophyte generation undergoes an asymmetrical rather than a symmetrical cell division and produces an upright filament and a rhizoid rather than the prostrate filament typical of wild type sporophytes (Peters et al., 2008). Individuals that carry this mutation therefore fail to implement the typical early sporophyte developmental program and resemble gametophytes but produce the sexual structures of the sporophyte generation at maturity. The absence of a phenotype during the gametophyte generation suggests that the developmental program directed by the *IMM* gene may have been a sporophyte-specific innovation.

Here we describe the positional cloning of the *IMM* locus and show that this gene encodes a protein of unknown function, which shares a novel, repeated motif with a viral protein. The *IMM* gene is part of a large, rapidly evolving gene family in *Ectocarpus* sp. and species with identifiable homologues exhibit an unusual distribution across the eukaryotic tree of life.

RESULTS

Positional cloning of the *IMM* locus

Peters *et al*. (2008) showed that the *imm* mutation behaved as a recessive, Mendelian allele and was located on an autosome. To map this mutation, a backcrossed descendant of strain Ec137 carrying the *imm* mutation (Peters et al., 2008) was crossed with the out-crossing line Ec568 (Heesch et al., 2010) to generate a segregating family of 1,699 haploid progeny. The *IMM* locus was then mapped genetically by scanning the genome for linked microsatellite markers and fine mapping the mutation. To scan for linked markers, a subset of the population (between 30 and 75 individuals) was genotyped for 97 microsatellite markers (Heesch et al., 2010) distributed at approximately 30 cM intervals along the length of the entire genetic map. Additional markers were then generated for a region on chromosome 27 that exhibited co-segregation with the Imm^{+} phenotype and these were tested against the entire population to fine map the *IMM* locus. Overall, a total of 121 markers were genotyped (Table S1), allowing the *IMM* locus to be mapped to a region of 43.7 kbp between coordinates 2,299,499-2,343,206 on chromosome 27 (Fig 1A).

To identify the *imm* mutation within the 43.7 kbp interval, this region was amplified as a series of PCR products and the pooled products sequenced on an Illumina HiSeq platform. As the Imm^{-} phenotype is the result of a spontaneous mutation that was not originally present in the parent strain Ec17, we reconstructed reference sequences for the two parental haplotypes of this region by sequencing equivalent PCR products amplified from eight wild type haploid siblings of the *imm* strain Ec137. Comparison of the sequence data for the eight siblings with that from the Ec137 strain allowed polymorphisms inherited from the diploid parent sporophyte to be distinguished from the causal mutation. Sanger resequencing was used to validate polymorphisms detected by the Illumina sequencing approach and to generate sequence data for several short regions that were not covered by the Illumina sequence data. This approach identified a two base pair deletion within exon five of gene Ec-27_002610 as the causal mutation of the Imm^{-} phenotype (Fig 1A). No other mutations were detected within the mapped interval.

The *IMM* gene encodes a protein of unknown function related to a protein encoded by a brown algal virus

The *IMM* gene (locus Ec-27_002610) is predicted to encode an 862 amino acid (91.8 kDa) protein, which consists of a long N-terminal domain that shares no similarity with other *Ectocarpus* sp. genes or genes from other species, plus a C-terminal domain which includes

five imperfect tandem repeats of a 38 amino acid cysteine-rich motif (C-X4-C-X16-C-X2-H-X12, Fig 1B). The 38 amino acid cysteine-rich motif is very similar to a cysteine-rich repeated motif found in the EsV-1-7 protein of the *Ectocarpus* virus EsV-1 (Delaroque et al., 2001), which also contains five of these repeated motifs. Based on this similarity, hereafter we will refer to the 38 amino acid cysteine-rich motif as an EsV-1-7 repeat.

The two base pair deletion in the *imm* mutant causes a frame shift in the part of the gene that encodes the N-terminal domain. The mutation is predicted to lead to the production of a 418 amino acid protein with a truncated N-terminal domain and possessing no EsV-1-7 repeats (Fig 1B).

Disruption of *IMM* function using RNA interference

Recent work has demonstrated that injection of double stranded RNA into zygotes of the brown alga *Fucus* induces an RNA interference response, leading to knockdown of target gene expression (Farnham et al., 2013). RNA interference therefore represents a potential approach to investigate gene function in brown algae but a modification to the *Fucus* protocol was required because microinjection is not feasible for *Ectocarpus* due to the small size of its cells. We therefore developed an alternative approach in which synthetic siRNA molecules were introduced into naked gametes using a transfection reagent (see Materials and Methods for details).

Wild type gametes that fail to fuse with a gamete of the opposite sex can develop parthenogenetically to give rise to partheno-sporophytes. These partheno-sporophytes go through the same developmental steps as diploid sporophytes derived from zygotes and are morphologically indistinguishable from the latter. In both cases the initial cell undergoes a symmetrical division that gives rise to two germ tubes, which grow to form a symmetrical, prostrate basal filament (Peters et al., 2008). The basal filament, which is composed of characteristic round and elongated cells, adheres strongly to the substratum. Following simultaneous introduction of three siRNA molecules targeting the *IMM* gene transcript, a small proportion (0.63% ± 0.09% mean ± standard deviation in six replicates each of 400 individuals) of the parthenogenetic gametes adopted a pattern of early development that closely resembled the phenotype of the *imm* mutant (Peters et al., 2008). These gametes underwent an asymmetrical rather than a symmetrical initial cell division and the two germ tubes of the developing partheno-sporophyte gave rise to an upright filament and a rhizoid (Fig 2). Individuals with this phenotype were not observed in parallel samples of gametes treated with an siRNA directed against a green fluorescent protein gene sequence as a control

(six replicates of 400 individuals) and the difference between the test and control experiments was was highly significant (Pearson's χ^2 = 13.0, p = 0.0003). These observations indicated that RNA-interference-induced knockdown of *IMM* gene expression had the same developmental consequences as the *imm* mutation, at least in a small proportion of the treated individuals. Taken together with the results of the mapping of the genetic mutation, this observation confirmed that the two base pair deletion identified in exon five of gene Ec-27_002610 was the causal mutation of the Imm$^-$ phenotype.

Expression pattern of the *IMM* gene

Quantitative reverse transcriptase PCR analysis indicated that the *IMM* gene transcript was approximately twice as abundant in diploid sporophytes and partheno-sporophytes as in gametophytes (Fig 3). The relatively high abundance of the *IMM* transcript during the gametophyte generation was surprising because no visible phenotype was detected during this generation in the *imm* mutant (Peters et al., 2008). The transcript was less abundant in *imm* mutant partheno-sporophytes than in wild type partheno-sporophytes (Fig 3), suggesting that the mutation has a destabilising effect on the transcript.

Analysis of gene expression in the *imm* mutant sporophyte

In a previous study we analysed gene expression in the *imm* mutant using a microarray constructed with sequences from two subtraction libraries enriched in genes differentially expressed during either the sporophyte or the gametophyte generation (Peters et al., 2008). This analysis indicated that sporophyte-specific genes were down regulated and gametophyte-specific genes upregulated in the *imm* mutant during the sporophyte generation. Based on this information, and the morphological resemblance of the *imm* sporophyte to the wild type gametophyte, the Imm$^-$ phenotype was interpreted as representing partial homeotic switching from the sporophyte to the gametophyte developmental program (Peters et al., 2008). Here we used multiple RNA-seq datasets to compare the *imm* transcriptome with a broader range of samples, including two microdissected partheno-sporophyte tissue samples corresponding to the apical upright filaments and the basal system, respectively. Principal component analysis indicated that, overall, the transcriptome of the *imm* partheno-sporophyte was actually more similar to the transcriptomes of wild type partheno-sporophyte samples, particularly samples that included upright filaments, than to wild type gametophyte samples (Fig 4).

We therefore reanalysed the expression patterns of sets of genes that had been previously identified as being significantly upregulated or down regulated in the *imm* mutant partheno-

sporophyte compared to the wild type partheno-sporophyte (denoted as *imm* upregulated or IUP and *imm* downregulated or IDW genes by Peters et al., 2008). This analysis indicated that IUP and IDW genes tended to be up and down regulated, respectively, in gametophyte samples but they also showed very similar patterns of expression in upright filaments isolated from the sporophyte generation (Fig S1). Bearing in mind that that the gametophyte generation consists almost entirely of upright filaments, these results suggested that the IUP and IDW genes, rather than being life-cycle-regulated genes, may correspond to loci that are differentially regulated in upright filaments compared to basal tissues.

To further investigate this possibility, a genome-wide analysis was carried out using RNA-seq data and DEseq2 (Love et al., 2014) to identify additional genes that were differentially expressed in the *imm* partheno-sporophyte compared to the wild type partheno-sporophyte. This analysis identified 1,578 genes that were significantly differentially expressed between the two samples (1,087 up-regulated in *imm* and 491 down-regulated; Table S2). Again, analysis of expression patterns across several different samples indicated that the majority of these genes did not exhibit life-cycle-generation-specific expression patterns but rather were up- or down-regulated in upright filaments (Fig S2).

Taken together, these observations suggest an alternative interpretation of the Imm$^-$ phenotype. Rather than representing a mutation that causes switching between life cycle generations, we propose that abrogation of the *IMM* gene leads to failure to correctly implement the early developmental program of the sporophyte. In the absence of a functional *IMM* gene the initial cell does not divide symmetrically and there is no deployment of a system of basal filaments before the establishment of the apical/basal axis (Fig 2A). Rather, an asymmetrical division of the initial cell directly produces a basal rhizoid cell and an apical thallus cell. We suggest that the resemblance with the gametophyte, in terms of gene expression, is not due to switching to the gametophyte developmental program but rather due to the sporophyte adopting an alternative developmental program which is more similar to that of the gametophyte (immediate production of an upright filament).

The 1,578 genes that were identified as being significantly differentially expressed between the *imm* partheno-sporophyte and the wild type partheno-sporophyte were analysed for enriched gene ontology categories. One significant (FDR < 5%) category was found for the upregulated genes (G-protein signalling) and two (photosynthesis-related and RNA polymerase II activity) for the downregulated genes (Tables S3 and S4).

Figures

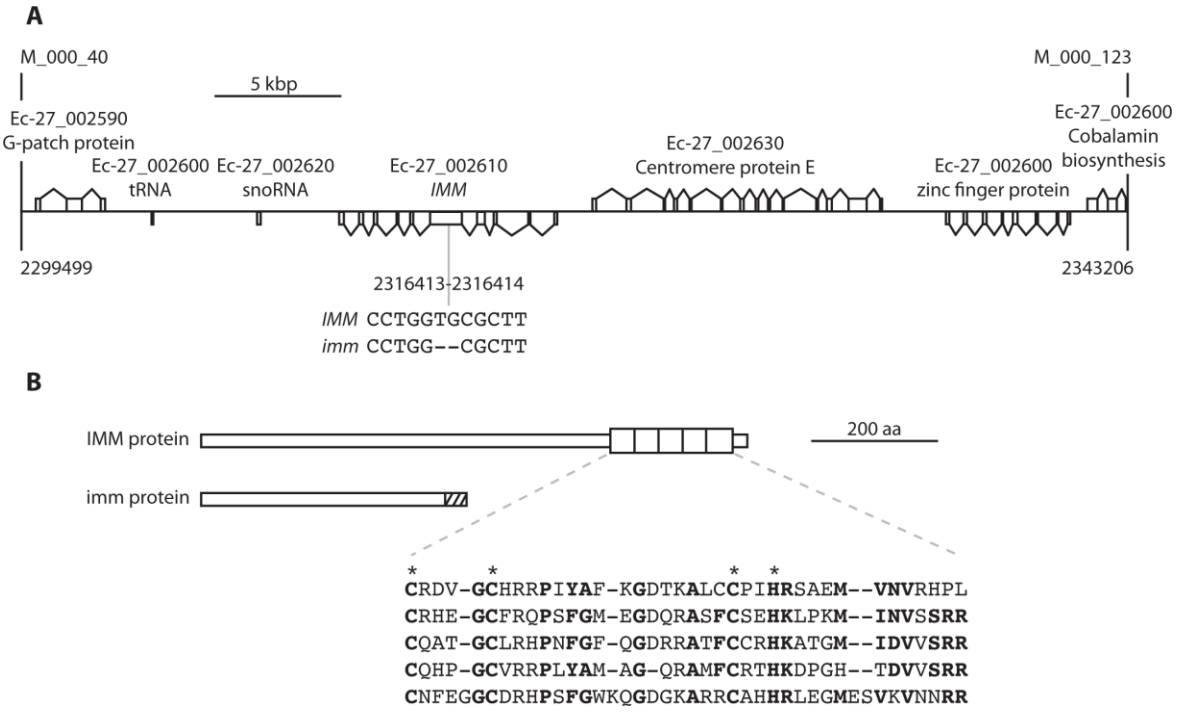

Fig. 1. **Positional cloning of the *IMM* gene and features of the encoded protein.**

(A) Schematic representation of the 43,708 bp interval on chromosome 27 between the closest recombining markers to the *IMM* locus. Protein coding exons, a tRNA and a snoRNA locus are shown as boxes. Genes above the line are transcribed to the right, genes below the line to the left. The position of the single mutation within the mapped interval (within exon five of the gene Ec-27_002610) is indicated. (B) Schematic representation of the wild type IMM protein and the predicted product of the *imm* mutant allele including an alignment of the five EsV-1-7 repeats at the C-terminal end of the IMM protein. Conserved residues are shown in bold and four highly conserved cysteine and histidine residues are indicated with an asterisk.

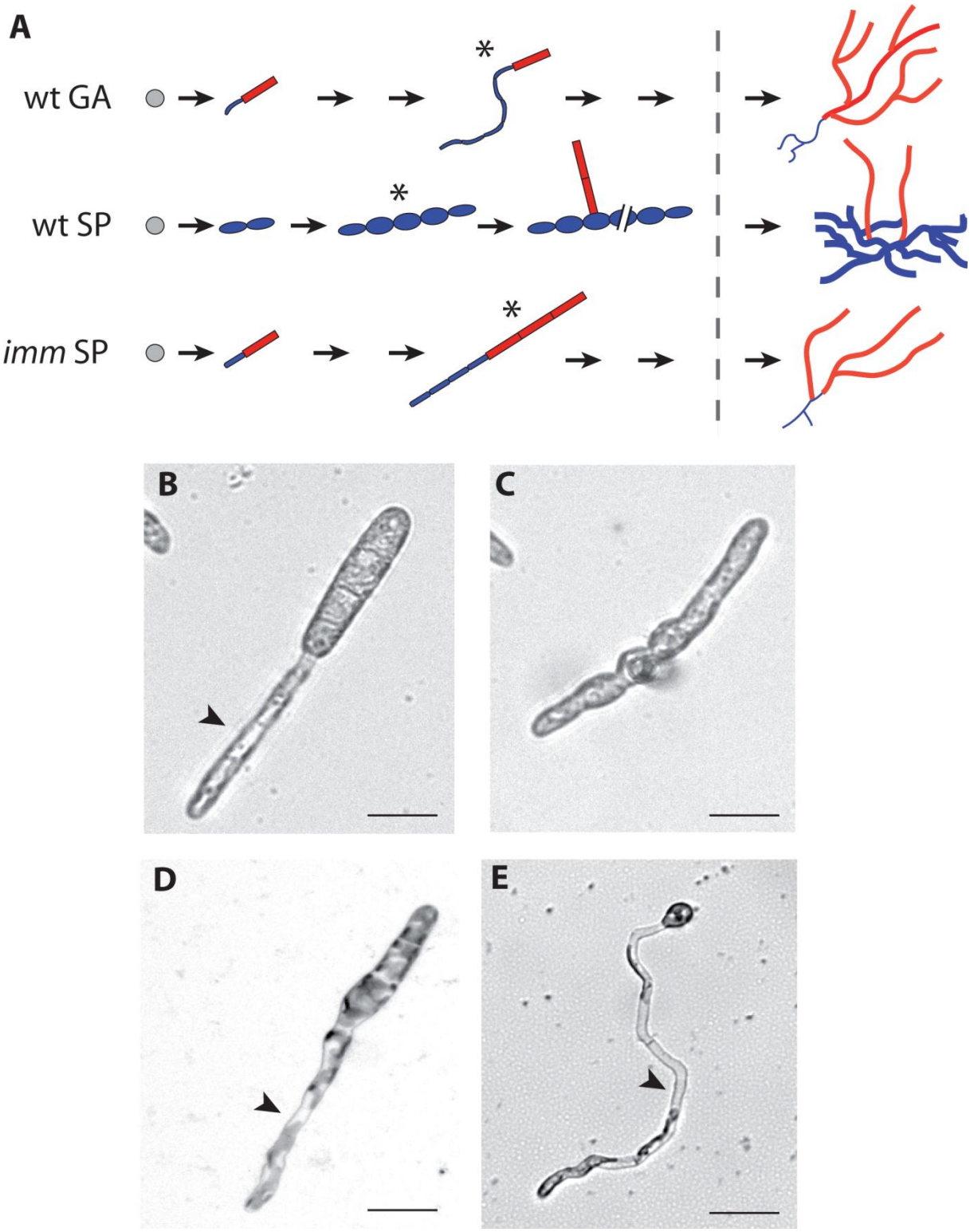

Fig. 2. Introduction of siRNAs that target the *IMM* gene induces a phenotype that mimics the *imm* mutation.