

## ***Alternative Splicing in the Obligate Biotrophic Oomycete Pathogen *Pseudoperonospora cubensis****

The Faculty of Oregon State University has made this article openly available.  
Please share how this access benefits you. Your story matters.

<b>Citation</b>	Burkhardt, A., Buchanan, A., Cumbie, J. S., Savory, E. A., Chang, J. H., & Day, B. (2015). Alternative splicing in the obligate biotrophic oomycete pathogen <i>Pseudoperonospora cubensis</i> . <i>Molecular Plant-Microbe Interactions</i> , 28(3), 298-309. doi:10.1094/MPMI-09-14-0300-FI
<b>DOI</b>	10.1094/MPMI-09-14-0300-FI
<b>Publisher</b>	American Phytopathological Society
<b>Version</b>	Version of Record
<b>Terms of Use</b>	<a href="http://cdss.library.oregonstate.edu/sa-termsfuse">http://cdss.library.oregonstate.edu/sa-termsfuse</a>

# Alternative Splicing in the Obligate Biotrophic Oomycete Pathogen *Pseudoperonospora cubensis*

Alyssa Burkhardt,<sup>1</sup> Alex Buchanan,<sup>2</sup> Jason S. Cumbie,<sup>2,3</sup> Elizabeth A. Savory,<sup>2</sup> Jeff H. Chang,<sup>2,3,4</sup> and Brad Day<sup>1,5,6</sup>

<sup>1</sup>Graduate Program in Cell and Molecular Biology, Michigan State University, East Lansing, MI 48824, U.S.A.; <sup>2</sup>Department of Botany and Plant Pathology, <sup>3</sup>Molecular and Cellular Biology Program, and <sup>4</sup>Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR 973314, U.S.A.; <sup>5</sup>Graduate Program in Genetics, and <sup>6</sup>Department of Plant, Soil and Microbial Sciences, Michigan State University

Submitted 30 September 2014. Accepted 30 October 2014.

*Pseudoperonospora cubensis* is an obligate pathogen and causative agent of cucurbit downy mildew. To help advance our understanding of the pathogenicity of *P. cubensis*, we used RNA-Seq to improve the quality of its reference genome sequence. We also characterized the RNA-Seq dataset to inventory transcript isoforms and infer alternative splicing during different stages of its development. Almost half of the original gene annotations were improved and nearly 4,000 previously unannotated genes were identified. We also demonstrated that approximately 24% of the expressed genome and nearly 55% of the intron-containing genes from *P. cubensis* had evidence for alternative splicing. Our analyses revealed that intron retention is the predominant alternative splicing type in *P. cubensis*, with alternative 5'- and alternative 3'-splice sites occurring at lower frequencies. Representatives of the newly identified genes and predicted alternatively spliced transcripts were experimentally validated. The results presented herein highlight the utility of RNA-Seq for improving draft genome annotations and, through this approach, we demonstrate that alternative splicing occurs more frequently than previously predicted. In total, the current study provides evidence that alternative splicing plays a key role in transcriptome regulation and proteome diversification in plant-pathogenic oomycetes.

*Pseudoperonospora cubensis* is the causal agent of cucurbit downy mildew, one of the most agriculturally important foliar diseases of the Cucurbitaceae family (Burkhardt and Day 2013). Like other oomycete members of the Stramenopiles, Heterokonts, and Alveolates (SAR) supergroup, *P. cubensis* is highly destructive, with epidemics of downy mildew on cucumber (*Cucumis sativus* L.) causing up to 100% yield loss

A. Burkhardt and A. Buchanan are co-first authors of this work.

Updated *Pseudoperonospora cubensis* genome annotation and alternative splicing predictions is archived at Oregon State University Libraries: [dx.doi.org/10.7267/N9TD9V7M](http://dx.doi.org/10.7267/N9TD9V7M).

Corresponding authors: J. H. Chang (computational queries); E-mail: [jeff.chang@science.oregonstate.edu](mailto:jeff.chang@science.oregonstate.edu); and B. Day (biological queries); E-mail: [bday@msu.edu](mailto:bday@msu.edu)

\*The e-Xtra logo stands for “electronic extra” and indicates that five supplementary figures and three supplementary tables are published online.

(Beakes et al. 2012; Burkhardt and Day 2013; Savory et al. 2010). In recent years, particularly in the United States, the effectiveness of host resistance has decreased and is no longer sufficient for management of the disease (Holmes et al. 2006). The economic losses, coupled with rising fungicide costs associated with controlling *P. cubensis* outbreaks, highlight the importance of studying the *P. cubensis*–cucumber pathosystem.

*P. cubensis* is a challenging pathogen to study due to its obligatory host-associated lifestyle and its intractability to genetic manipulation. The development of genomic and transcriptomic resources is difficult because of the need to disentangle pathogen sequences from those of the host. Some of these difficulties have been overcome, and draft genome and transcriptome reference sequences have been generated for both the pathogen and its cucumber host (Adhikari et al. 2012; Savory et al. 2012a and b; Tian et al. 2011). Furthermore, transcriptome-wide changes have been documented for both *P. cubensis* and cucumber during infection that characterized gene expression during different life stages of *P. cubensis*, including free-swimming sporangia, encystment, invasion via stomata, and hyphal expansion through a susceptible cucumber variety (Adhikari et al. 2012; Savory et al. 2012a).

The draft genome sequence of *P. cubensis* has also been mined for candidate virulence determinants, such as effector proteins that may play a role in *P. cubensis* pathogenesis (Savory et al. 2012b; Tian et al. 2011). Oomycetes rely on both apoplastic and cytoplasmic effector proteins, which are secreted into the apoplast or the extrahaustorial matrix, respectively, to facilitate infection of their hosts (Kamoun 2006). Candidate effectors can be predicted based on the presence of a functional signal peptide, which is a feature necessary for secretion from the pathogen. Additionally, cytoplasmic effectors that are translocated across the host membrane can be identified based on the presence of a short RXLR amino acid motif (Rehmany et al. 2005; Whisson et al. 2007) or, as shown in *P. cubensis*, an alternative QXLR motif (Tian et al. 2011). Based on the original draft genome sequence, *P. cubensis* is predicted to encode 271 secreted effector proteins, of which 67 had identifiable R- or QXLR motifs, that were found to be expressed at all stages of infection surveyed (Savory et al. 2012a and b; Tian et al. 2011). The first *P. cubensis* effector identified, *PcQNE*, is a protein that contains a QXLR motif that is necessary for translocation into host cells (Tian et al. 2011). The second characterized effector protein from *P. cubensis*, *PscRXLR1*, is unique in that its gene has evidence for alternative splicing (AS). The product of the constitutively spliced gene product was predicted to encode a multidrug transporter

which, in many analyses, would likely be dismissed as a false positive. The alternatively spliced transcript, on the other hand, is predicted to yield a candidate-secreted RXLR effector due to the retention of an intron and the introduction of a premature termination codon (Savory et al. 2012b).

AS is a fundamental and evolutionarily conserved eukaryotic process that leads to multiple transcripts from a single gene (Chen et al. 2012; McGuire et al. 2008). In higher eukaryotes, AS plays important roles in development and stress signaling (Mastrangelo et al. 2012), circadian-dependent regulation of host physiology (James et al. 2012; Sanchez et al. 2011; Schöning et al. 2007), the modulation of gene expression in human genetic diseases (Nicholson et al. 2009; Pistoni et al. 2010), and the response of plants to pathogen infection (Rayson et al. 2012). Beyond diversifying the products of expressed genes, AS also has a role in regulating gene expression (Boothby et al. 2013; Guan et al. 2013). In microbial eukaryotes, AS is less characterized but has nevertheless been identified as an important mechanism for regulating gene expression and influencing the ability of eukaryotic microbes to infect their plant hosts (Juneau et al. 2007; Klopffholz et al. 2011; Pleiss et al. 2007; Savory et al. 2012b; Shen et al. 2011; Wilhelm et al. 2008).

Eukaryotic genes are transcribed as pre-mRNAs and, during their maturation, introns are spliced out and exons are joined via a process that is carried out by the spliceosome (Hoskins and Moore 2012). The intron–exon boundaries of transcripts are defined via a small number of short and degenerate *cis*-regulatory sequences, the 5' and 3' splice sites (5'ss and 3'ss), the branch point sequence, and the polypyrimidine tract (Keren et al. 2010). Additional degenerate enhancer and silencing sequences have also been shown to recruit splicing regulatory proteins to influence both constitutive splicing and AS (Keren et al. 2010). In the case of the latter, alternative splice sites are used to generate varying combinations of transcripts from a single gene locus, thereby regulating both the transcriptome and proteome (Kalsotra and Cooper 2011).

Two mechanisms of splice site selection have been proposed (Keren et al. 2010). In the first, exon definition, the spliceosome recognizes and places the initial machinery across the exon. Generally, this mechanism is more prevalent in higher eukaryotes (Keren et al. 2010). In the second, intron definition, the spliceosome is placed across introns and recognizes splice sites at each end of the same intron. This mechanism is predicted to be more ancient and common to eukaryotic microbes (Keren et al. 2010). The differences in mechanisms impose constraints that are reflected in gene architecture and in the degree and type of AS. Higher metazoans, such as mice and humans, tend to have short exons and long introns, with a high degree of AS and a preference for exon skipping (Kim et al. 2008; Pan et al. 2008; Sultan et al. 2008). Conversely, eukaryotic microbes such as yeast have shorter introns and lower frequencies of AS that favor intron retention (Irimia et al. 2007; Kim et al. 2008; Meyer and Vildardell 2008; Wilhelm et al. 2008). It is hypothesized that plants use intron definition and, in this regard, are more similar to eukaryotic microbes, with one notable exception being the high frequency of AS in plants compared with microbes (Filichkin et al. 2010; Marquez et al. 2012; Reddy 2007; Syed et al. 2012; Wang and Brendel 2006).

There has been limited analysis of AS in oomycetes. Previous analyses of *Phytophthora infestans* and *P. sojae* have relied on expressed sequence tag (EST) datasets, which could contribute to an underestimate of the true breadth of AS in this group of organisms (McGuire et al. 2008; Shen et al. 2011). For example, only 405 and 66 splice variants were identified in *P. infestans* and *P. sojae*, respectively, with intron retention being predominant in both species (McGuire et al. 2008). In a

more detailed study of *P. sojae*, unique sequences at the 5'ss were identified and it was suggested that these sequences may function in AS (Shen et al. 2011). However, of the 1,681 predicted intron-bearing gene models, only 122 were identified as having an AS event (Shen et al. 2011). These observations of a low frequency of AS are in conflict with the low strength of the 5'ss of introns identified in *P. sojae* because low strength is typically an indicator of intron richness and prominent AS (Irimia et al. 2007; Shen et al. 2011). Additionally, members of the SAR group have introns in 35 to 86% of their genes, indicating a potential for higher rates of splicing and AS than predicted from analysis of ESTs (Supplementary Fig. S1).

Based on its original genome reference sequence, *Pseudoperonospora cubensis* has the lowest percentage of intron-bearing genes (approximately 35%) among the oomycetes surveyed. *P. cubensis* averages 0.7 introns per gene, with the average intron length being 86 nucleotides (nt). In the current study, we used RNA-Seq to identify expressed genes from *P. cubensis*, and used this information to update existing gene models, to identify new genes, and to predict AS events; the latter was prompted by a previous study from our group that identified and characterized an alternatively spliced effector-encoding gene (Savory et al. 2012b). Herein, we show that, of intron-bearing genes (24% of the expressed genes), approximately 55% exhibited AS, including alternative 3', alternative 5', and intron retention events. Representatives of the newly identified genes and predicted alternatively spliced transcripts were validated for expression and AS. We suggest that AS in oomycete plant pathogens is substantially higher than previously predicted and is at a level closer to that of plants. Furthermore, AS was observed in genes predicted to encode signal peptide-containing proteins and putative effectors, which provides support for the hypothesis that AS has the potential to influence the virulence of this important plant pathogen.

## RESULTS

### Generation and analysis of *P. cubensis* RNA-Seq datasets.

We used RNA-Seq to resequence cDNA libraries derived from *P. cubensis* during infection of its cucumber host to quantify transcripts with evidence for AS. We used paired-end (PE) sequencing on an Illumina HiSeq and generated a total of approximately 735 million cDNA fragments (170 to 210 million fragments per sample) from bar-coded RNA-Seq libraries derived from two biologically replicated samples of 2, 3, 4, and 8 days postinoculation (dpi) (Adhikari et al. 2012; Savory et al. 2012a). The PE RNA-Seq reads from the greater than 20.5 million previously sequenced fragments of *P. cubensis* sporangia (Savory et al. 2012a) were also included in this study.

More than 750 million sequenced fragments were preprocessed for quality and categorized based on origin. Consistent with the previous study, the majority of the reads mapped to the host (Savory et al. 2012a). Only approximately 40 million sequenced fragments aligned uniquely to and within the *P. cubensis* reference sequence, either as an ungapped alignment with up to two mismatches or as a gapped alignment with no mismatches. Fragments from sporangia, intermediate (2 to 4 dpi), and late stages (8 dpi) of the infection composed approximately 40, 30, and 30%, respectively, of the total aligned fragment sequences. The majority of the greater than 750 million sequenced fragments (approximately 650 million sequenced fragments) aligned with greater confidence to the *C. sativus* reference sequence. The remaining >60 million sequenced fragments did not align to either reference sequence. Approximately 75% of these sequences were likely replete with sequencing errors, because they were rare ( $\leq 10$  RNA-Seq reads). The other unaligned sequenced fragments likely correspond to

other microorganisms living in or on the infected cucumber leaf. The experiment was not conducted in a sterile environment and BLASTN analysis (threshold  $E$  value =  $1 \times 10^{-5}$ ) of 1,000 randomly selected read sequences used as queries returned homologies to sequences of prokaryotic origin.

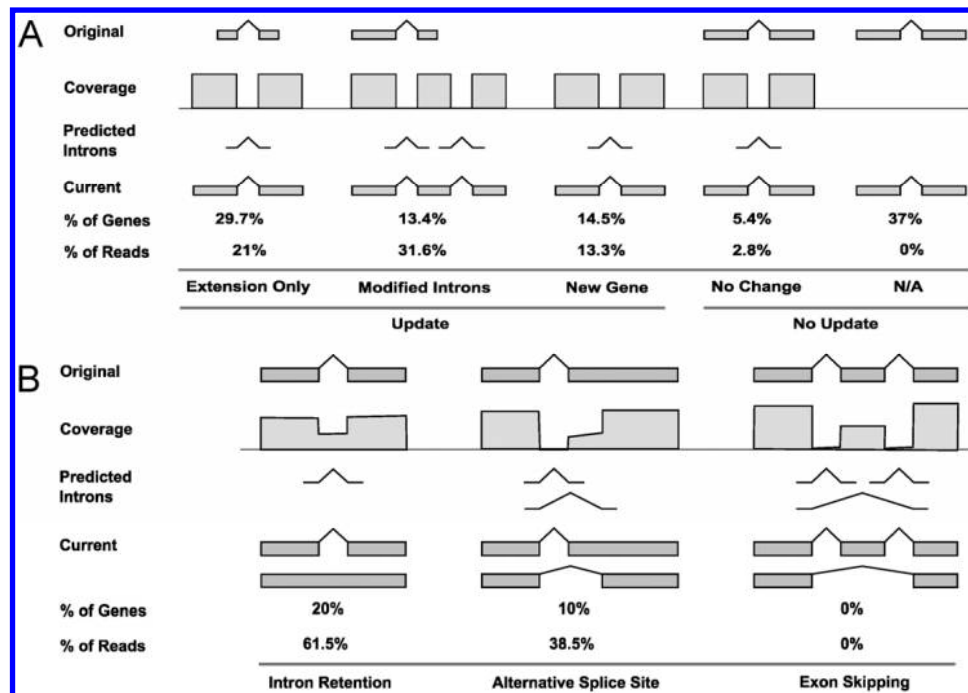
### Improvements to the *P. cubensis* expressed genome annotation.

In order to more confidently predict AS, we first used the RNA-Seq sequences to improve the annotation of the *P. cubensis* reference genome sequence (Fig. 1A). Nearly 45% of the reference genome sequence was aligned, with an average depth of 218 RNA-Seq reads per sequenced nucleotide. In total, approximately 30 million sequenced fragments, aggregated across all libraries, aligned to 13,483 of the 23,519 (57%) previously annotated genes and were classified as expressed (Savory et al. 2012a). Another 8.2 million sequenced fragments aligned to regions of the genome devoid of annotated features. Those that coalesced into distinct clusters exceeding length and coverage thresholds were used to improve the genome annotation. The majority of the clusters of reads were used to improve 11,892 genes annotated in the original genome reference sequence. Of the updated genes, approximately 29.7% were extended and another 13.4% had intron definitions altered; some genes were both extended and had intron definitions altered. Among the extended genes, approximately 2,000 of the improvements were the addition of new coding sequences, and 10,779 were addition of 5' or 3' untranslated regions (UTR). In addition to the alterations to loci of the original genome annotation, we also classified 4,072 features as new candidate gene loci (14.5%), thus increasing the expressed genome of *P. cubensis* to approximately 17,500 genes (approximately 64% of the total number of annotated genes). A small set of genes was selected based on the predicted level of expression in sporangia, and was tested using reverse-transcrip-

tase polymerase chain reaction (RT-PCR) to confirm our prediction of new genes (Supplementary Fig. S2). From this point on, we refer only to the expressed genome, which consists of 17,558 expressed genes identified based on analysis of the RNA-Seq datasets.

Approximately 2 million sequenced fragments uniquely mapped as perfectly but gapped alignments to the *P. cubensis* reference genome sequence. Using this output, we generated a list of approximately 20,000 putatively spliced sequences. Of these, 9,964 corresponded to an exon-exon junction sequence present in 6,366 expressed intron-bearing genes, which represents 65% of all possible splice junctions in the expressed transcriptome. Additionally, the majority of the predicted spliced sequences were supported by a high number of RNA-Seq reads. In 169 cases, reads corresponding to putatively spliced sequences aligned within annotated exons in which the intervening regions had sequence coverage that fell below operational thresholds for expression. Based on this, we updated the annotations of the regions to introns. The converse situation was also observed, in which 700 annotated intronic sequences had high RNA-Seq coverage but no gap-aligned reads that spanned the annotated introns. These were reannotated as exons. Improvements to the genome annotation did not significantly change the average length of introns.

We also identified genes in the expressed genome that had signatures associated with genes involved in oomycete pathogenesis. We first analyzed translated sequences for predicted signal peptides. We predicted 1,736 and 2,139 secreted proteins from the translated sequences of the original annotated and updated expressed genomes, respectively. Of these, 1,320 were shared in both datasets. In the updated expressed genome, we identified an additional 819 genes predicted to encode secreted proteins that had not been annotated previously. In addition, 416 proteins predicted to be secreted based on the original annotation were not identified in the expressed genome.



**Fig. 1.** *Pseudoperonospora cubensis* genome annotation improvement using RNA-Seq. **A**, Genes were first improved based on the coverage of sequencing reads to the reference sequence. **B**, Gene models were predicted based on the alignment of reads to intron sequences or those that aligned as gapped reads. Original: examples of gene models from the original genome annotation. Coverage: example of RNA-Seq read coverage. Predicted Introns: based on original annotation or gap alignment of RNA-Seq reads. Current: improved gene or gene model annotation. Percentage of genes: ratio of the number of genes affected in each category, divided by the total number of original genes plus newly identified genes. Percentage of reads: ratio of the number of RNA-Seq reads used in each category, divided by the total number of mapped reads (for A, 31.3% of reads were not used); N/A = not applicable.

Because we used the same software version to mine the original annotation and the expressed genome for genes that encode putative secreted proteins, we concluded that differences are likely a consequence of improvements via extensions at the 5' end of gene models.

We next examined the putative secretome for proteins with RXLR and QXLR motifs (hereafter, we will simply refer to these proteins as RXLR). We identified 59 additional protein sequences that contained the RXLR translocation motif, increasing the number of putative RXLR-effectors from 67 to 125. Of these, 54 had been identified in the original genome annotation. Our improvements to the gene models resulted in the prediction of 59 new candidates; 13 previously annotated putative RXLR proteins were unsupported. In addition, the RNA-Seq data supports expression of 125 of these candidate RXLR effector proteins.

### Genome-wide identification of alternatively spliced genes in *P. cubensis*.

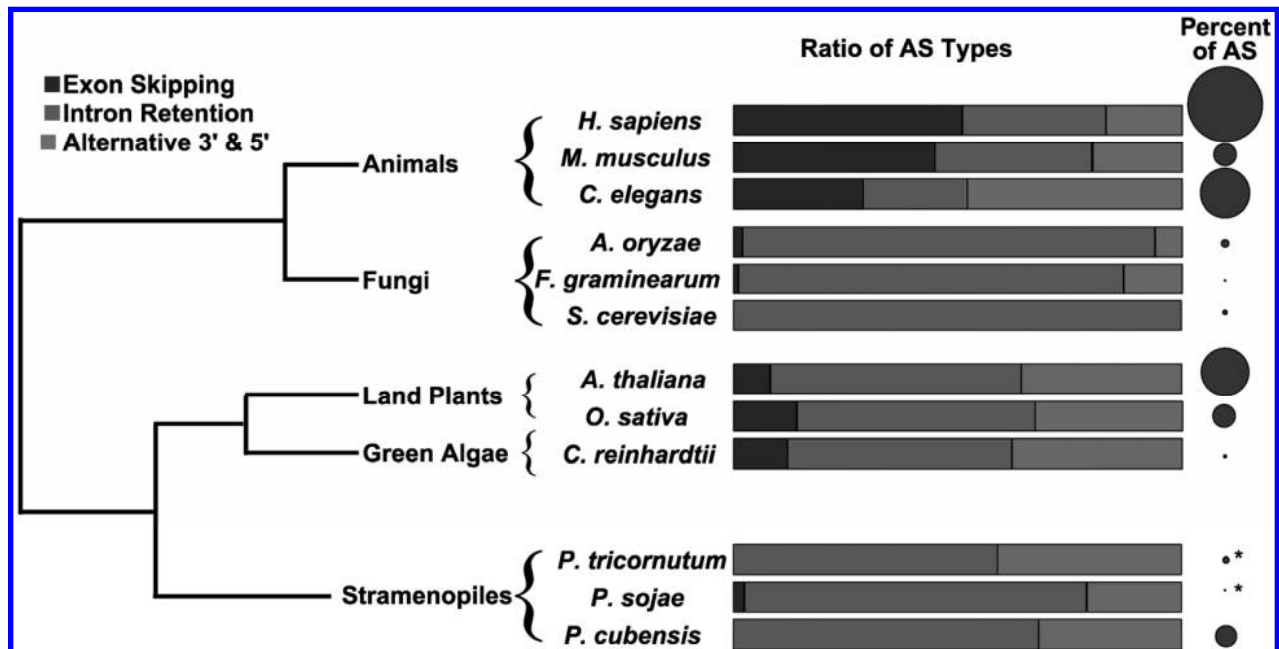
Using stringent criteria for the prediction of alternatively spliced transcript isoforms, we identified approximately 10,000 potential AS events associated with 4,205 genes (Fig. 1B; Table 1). Of these genes, 3,492, representing 20% of the expressed genome, had evidence for intron retention. In total, 1,812 genes (approximately 10%), many of which also had

evidence for intron retention, had evidence for either alternative 5'ss or 3'ss selection (Fig. 1B). Surprisingly, despite the use of multiple in silico approaches, we failed to identify any incidences of exon skipping. A total of 24% of the expressed genome and 55% of the intron-bearing genes had evidence for AS (Table 1).

Next, we projected AS preferences and levels onto a phylogenetic tree with representative members of key lineages, including fungi, stramenopiles, plants, and animals, that had publicly available genome or transcriptome data (Fig. 2; Supplementary Table S1). As a result of our improvements, the percent of intron-bearing genes in *P. cubensis* increased from 35% in the original annotation to 43% in the expressed genome, with an average of 0.86 per gene and 2.57 introns per intron-bearing gene. Among the members of the SAR group studied, *P. cubensis* has the fewest intron-bearing genes. Nevertheless, compared with the two members of the SAR group with available AS data available, *Phytophthora sojae* and *Phaeodactylum tricornutum*, *Pseudoperonospora cubensis* has a higher portion of alternatively spliced genes (Bowler et al. 2008; Shen et al. 2011). The preference in AS types and the overall level of AS predicted for *P. cubensis* were more similar to those of land plants than any other lineage investigated. Plants favor intron retention but do vary in levels of AS. Specifically, 61% of multiexonic *Arabidopsis thaliana* genes were spliced, compared with only 3% of those in the green algae *Chlamydomonas reinhardtii* (Labadorf et al. 2010; Marquez et al. 2012). In contrast, members of the animal lineage prefer exon skipping and can have extremely high levels of AS which, in *Homo sapiens*, is often tissue or cell specific (Keren et al. 2010; Pan et al. 2008). The fungi, with the exception of *Saccharomyces cerevisiae* and *S. pombe*, have high (>75%) percentages of intron-bearing genes. However, they all have very low numbers of alternatively spliced genes, most of which were intron retention. The highest is *Aspergillus oryzae*, with 11.10% of intron-bearing genes alternatively spliced (Wang et al. 2010).

**Table 1.** Statistics for the *Pseudoperonospora cubensis* transcriptome

Parameters	Number (%)
Total expressed genes	17,558
Constitutively spliced genes	13,353 (76.1)
Encode a predicted protein (not secreted)	11,576 (65.9)
Encode a predicted secreted protein	1,645 (9.4)
Encode a predicted effector	88 (0.5)
Alternatively spliced genes	4,205 (23.9)
Encode a predicted protein (not secreted)	3,674 (20.9)
Encode a predicted secreted protein	494 (2.8)
Encode a predicted effector	37 (0.2)



**Fig. 2.** Alternative splicing distribution across eukaryotic taxonomic branches. Representative members of the animal (*Homo sapiens* [95%], *Mus musculus* [28%], and *Caenorhabditis elegans* [62%]), fungi (*Aspergillus oryzae* [11.1%], *Fusarium graminearum* [1.7%], and *Saccharomyces cerevisiae* [6%]), plant (including land plants *Arabidopsis thaliana* [61%] and *Oryza sativa* [27.7%] and green algae *Chlamydomonas reinhardtii* [3%]), and Chromalveolata (Stramenopiles *Phaeodactylum tricornutum* [7.32%], *Phytophthora sojae* [2.29%], and *Pseudoperonospora cubensis* [24%]) kingdoms are included. Numbers in parentheses indicated the percentage of genes in each genome where AS events are found, represented by the circles in the figure (\* indicates enlarged, not to scale). Shaded bars represent the ratio of the total number of alternative splicing (AS) types found in the genomes of each species.

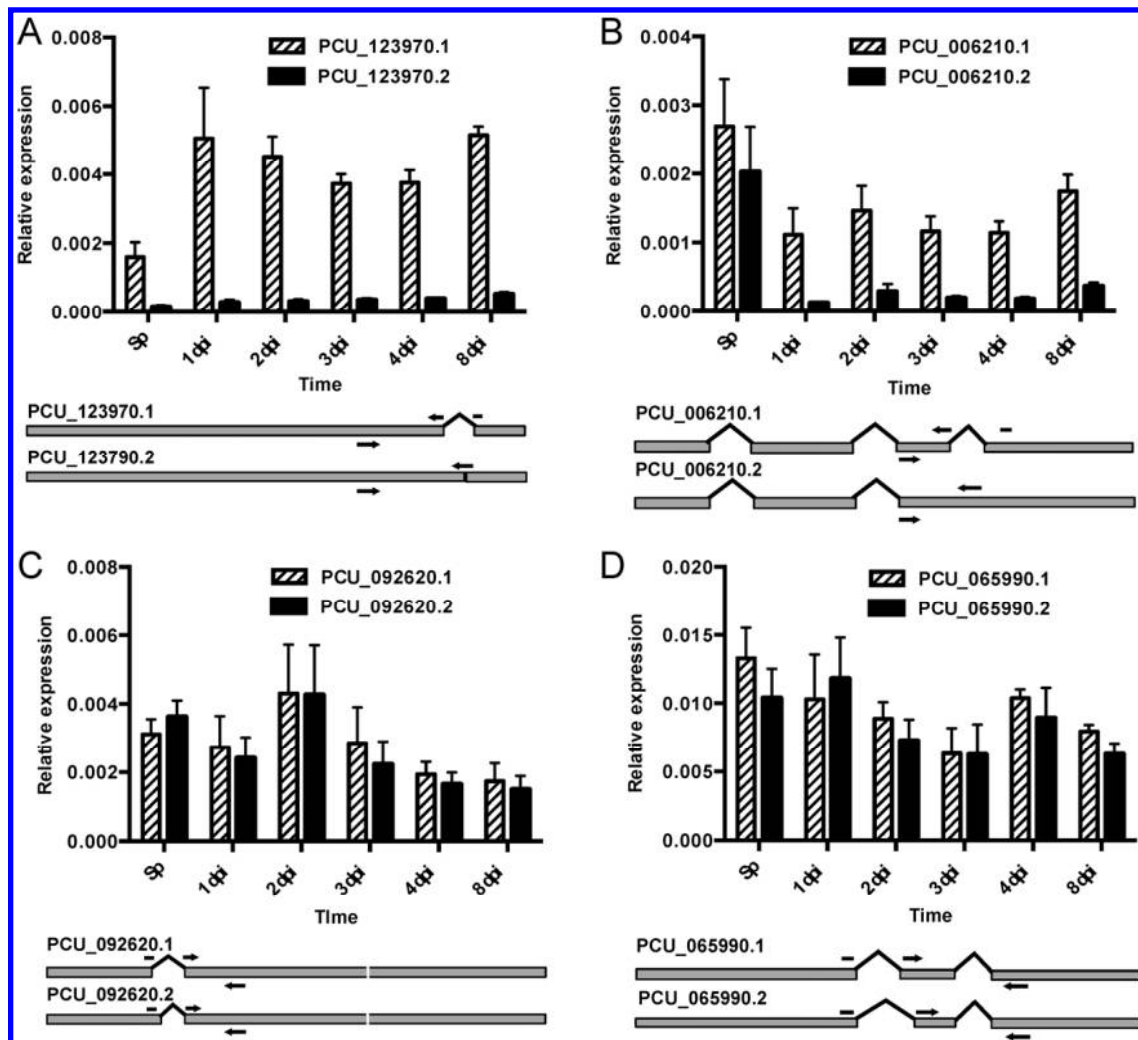


### Experimental validation of predicted AS events.

Because of the challenges associated with the small amount of biomass of an obligate biotrophic oomycete and its small intron sizes, we focused on a select number of highly expressed candidate genes that could be validated for AS using real-time PCR or RT-PCR. Candidate genes selected for validation were predicted to be expressed at every time point sampled and to have two isoforms. The gene models predicted by RNA-Seq data (Fig. 3; Supplementary Fig. S3) were validated using real-time PCR with isoform-specific primers. PCU\_123970.1 is predicted to encode a class IIb Aspartyl/Asparaginyl-tRNA synthetase (IPR002312) and lacks an identifiable signal peptide. RNA-Seq analysis suggested that the gene gives rise to an alternative intron-retaining isoform (PCU\_123970.2), which would introduce a premature termination codon and lead to a protein of only 65 amino acids. We could detect both the constitutively spliced and the intron-retaining isoforms during each of the developmental stages of the pathogen that were examined (Fig. 3A). However, there were clear differences in the expression patterns, consistent with inferences made from RNA-Seq analysis. PCU\_123970.1 showed higher relative expression than the intron-containing isoform, PCU\_123970.2, at all stages exam-

ined (unpaired *t* test *P* values for all pairwise comparisons at each time point  $\leq 0.03$ ). Furthermore, expression of PCU\_123970.1 increased substantially from sporangia to its host-associated stages (unpaired *t* test *P* values  $\leq 0.02$  for each pairwise comparison of sporangia to 2 to 8 dpi time points).

A second gene (PCU\_006210) that was tested also produces an alternative isoform that contains a retained intron (Fig. 3B). The translated product of PCU\_006210.1 is a 956-amino-acid protein with homology to a hypothetical protein of *Phytophthora parasitica* (GenBank: ETM43679.1). PCU\_0006210.2 is predicted to encode a 313-amino-acid product. The translated sequences for both isoforms are predicted to contain multiple breast cancer carboxy-terminal domains (peptide- and phosphopeptide-binding modules present in proteins involved in DNA checkpoint controls and DNA repair; IPR001357). Similar to the previous example, real-time PCR successfully confirmed the presence of both isoforms (Fig. 3B). In contrast, PCU\_006210.2 is expressed in sporangia to relative levels comparable with that of PCU\_006210.1 (unpaired *t* test *P* value = 0.521). The expression of both isoforms decreased in *Pseudoperonospora cubensis* during its host-associated stages, with PCU\_006210.2 expression dropping to barely detectable levels.

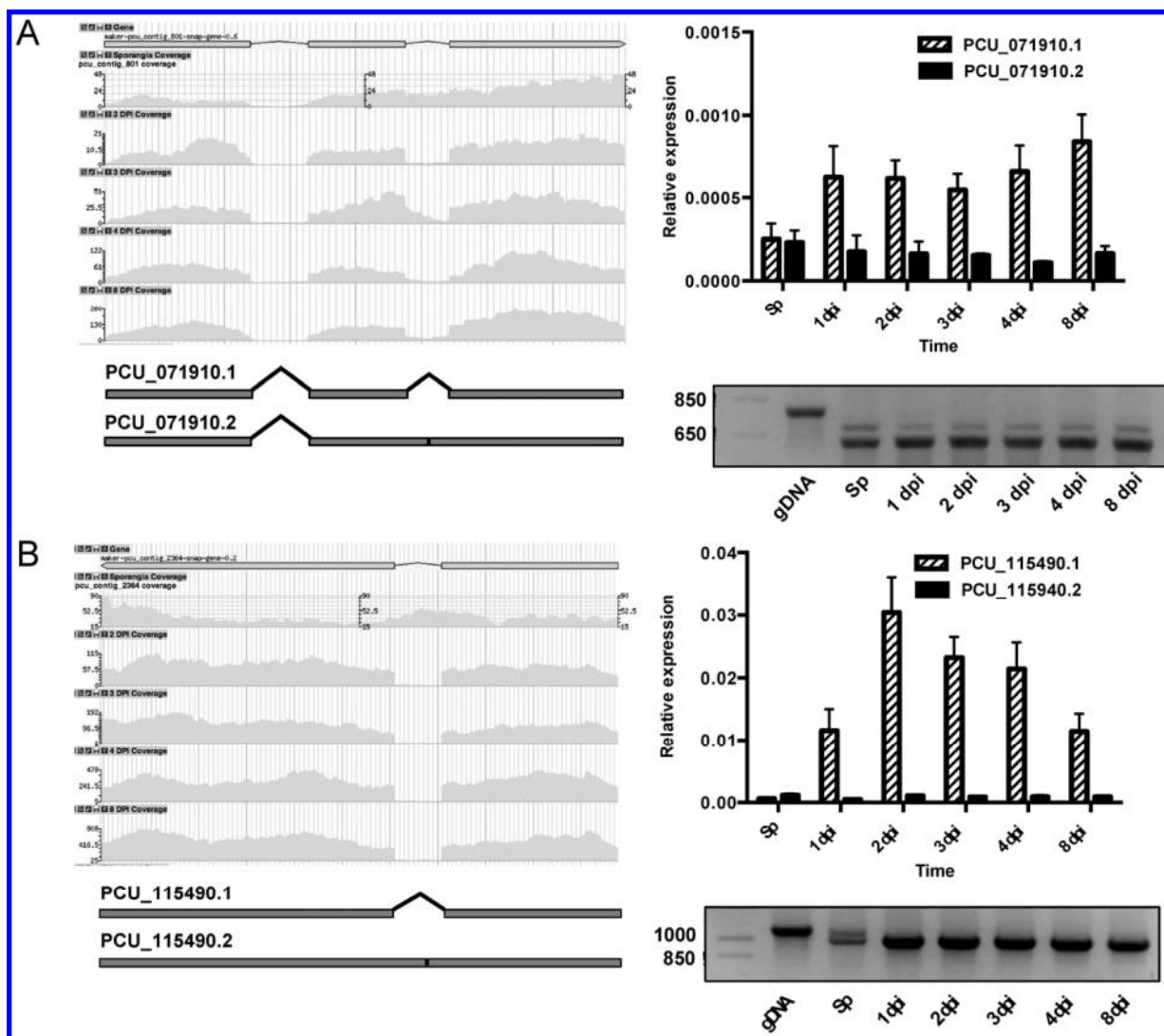


**Fig. 3.** Expression patterns of transcript isoforms of noneffector encoding genes exhibit different levels of relative expression over the course of infection. Real-time polymerase chain reaction (PCR) using isoform-specific primers (indicated by arrows on the gene model) was used to quantify the expression of each transcript isoform relative to the *Pseudoperonospora cubensis* internal control of the internal transcribed space region. **A**, Alternative splicing (AS) via the mechanism of intron retention in PCU\_123970. The vertical bar in the gene model for PCU\_123970.2 represents a premature stop codon in the predicted translated product. **B**, AS via intron retention in PCU\_006210. **C**, AS via the use of an alternative 5' splice site (ss) in PCU\_092620. **D**, AS via the use of an alternative 3' ss in PCU\_065990. Real-time PCR is shown as the average with standard error of three biological replicates; dpi = days postinoculation and Sp = sporangia.

We also used real-time PCR to confirm predicted alternative 5' splice sites (5'ss) and 3' splice sites (3'ss) usage. PCU\_92620 has a predicted N-terminal lipin domain (IPR007651) and is similar to a predicted lipin-like protein from *Phytophthora infestans* T30-4 (XP\_002903839.1). Both the fully spliced PCU\_92620.1 and its 5' alternatively spliced isoform PCU\_92620.2 appear to encode functional proteins of 649 and 660 amino acids, respectively. The two isoforms of PCU\_92620 were relatively highly expressed in all stages we examined, with both showing highest levels of expression at 2 dpi (Fig. 3C). PCU\_065990.1 is predicted to encode a member of the Rab5 family GTPase superfamily that is 200 amino acids long (IPR003579; GO:0005525), which functions in endocytosis and early endosome fusion (Schwartz et al. 2008). The second isoform is alternatively spliced via a 3' alternative splice site and, because of a frame shift, gives rise to a putative 91-amino-acid protein. Both isoforms of PCU\_065990 also showed relatively high expression throughout all stages tested but steadily decreased in expression over the course of sampling (Fig. 3D).

In all, 12% of the genes with evidence for AS are predicted to encode a secreted protein (Table 1). PCU\_071910.1 is pre-

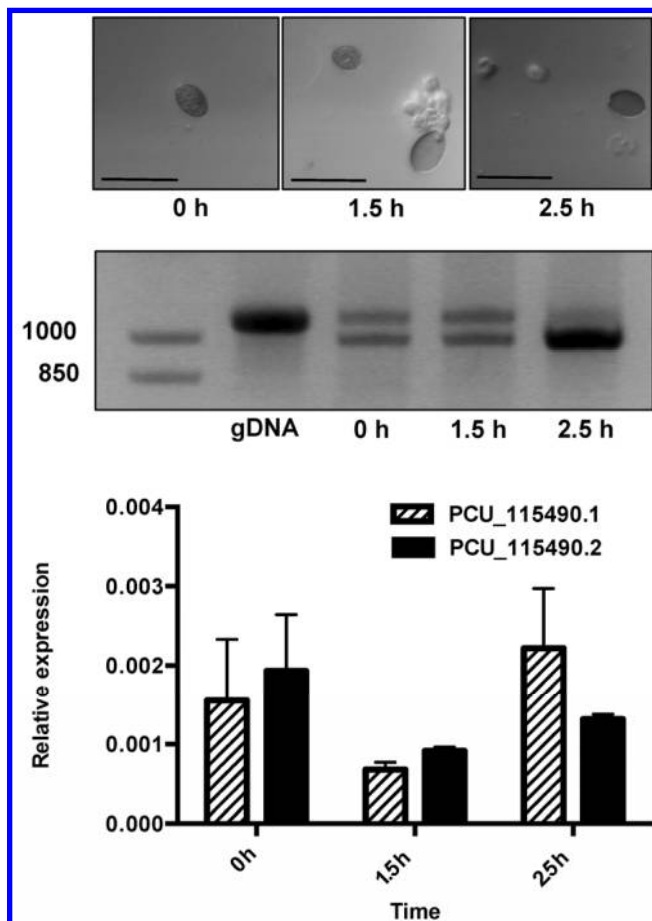
dicted to encode a secreted glutathione peroxidase (IPR000889; GO:0004602). The gene has two introns and, based on RNA-Seq coverage, we predict that PCU\_071910.2, which retains the second intron, will result in a truncated protein 141 amino acids long. The relative expression levels of PCU\_071910.1 and PCU\_071910.2 were similar in sporangia (Fig. 4A). During host infection, the PCU\_071910.1 isoform was significantly more expressed at 8 dpi relative to sporangia ( $t$  test  $P$  value = 0.035) and PCU\_071910.1 was significantly more expressed ( $t$  test  $P$  value < 0.05) relative to PCU\_071910.2 levels from 2 to 8 dpi (Fig. 4A). These data are consistent with the RNA-Seq results, which showed a similar decrease in the coverage of reads in the intron in samples collected during host infection relative to the sample collected from sporangia (Fig. 4A). Using Sanger sequencing, we confirmed the sequences for both transcript isoforms. To test whether AS affected localization of PCU\_071910, both isoforms were fused to cyan fluorescent protein (CFP) and expressed transiently in *Nicotiana benthamiana*. We did not observe any change in the localization of the protein isoforms (Supplementary Fig. S4).



**Fig. 4.** Genes encoding secreted gene products displaying alternatively splicing via intron retention. RNA-Seq coverage of each gene is shown as visualized through GBrowse. Coverage is shown for sporangia, 2, 3, 4, and 8 days postinoculation (dpi). Gene models are provided for each isoform, with a vertical bar indicating the location of a stop codon. Real-time polymerase chain reaction (PCR) was used to determine the relative expression of alternatively spliced isoforms as normalized to the *Pseudoperonospora cubensis* internal transcribed spacer region. Data are reported as the average relative expression and standard error from three biological replicates. Isoform products amplified from reverse-transcriptase (RT)-PCR were validated by Sanger sequencing and separated on a 1.5% agarose gel with a 1-kb Plus DNA ladder. Genomic *P. cubensis* DNA is shown as a positive PCR control. A, PCU\_071910; B, PCU\_115490. Sp = sporangia and DNA = PCR of genomic DNA from sporangia.

PCU\_115490 has a single intron and the translated sequence of the spliced form is predicted to encode a member of the endonuclease/exonucleases/peroxidase family (IPR005135) (Fig. 4B). Use of RT-PCR and real-time PCR confirmed the presence of an alternative intron-retaining form, which is predicted to lack all of the functional domains but retain its signal peptide sequence. Over the course of infection, PCU\_115490.1 shows a rapid increase in expression, peaking at 2 dpi, and gradually declining over the course of infection (Fig. 4B). The intron-retaining form, PCU\_115490.2, shows a consistent low level of expression. Thus, there is a dramatic change in the ratio of the predicted functional to putative nonfunctional isoforms for this gene, because PCU\_115490.1 consists of approximately 35% of the total PCU\_115490 gene expression in the sporangia and over 95% of the total expression by 1 dpi (Fig. 4B). We also fused CFP to both isoforms of PCU\_115490 and expressed transiently the genes in *N. benthamiana*. Similar to previous observations, changes in the localization of their protein isoforms were not observed.

The transition from sporangia to zoospore may be a key development change that correlates with changes in AS patterns.



**Fig. 5.** PCU\_115490 shows differential alternative splicing during pathogen developmental transitions from sporangia to zoospore. Representative developmental stages are shown in images collected over the time course. Images were taken using an Olympus IX71 inverted light microscope. Scale bars represent 200  $\mu$ m. Alternative isoforms were amplified using reverse transcriptase polymerase chain reaction (RT-PCR) with *Pseudoperonospora cubensis* RNA collected at 0, 1.5, and 2.5 h after exposure of sporangia to water. gDNA = PCR of genomic DNA from sporangia, 0 h = sporangia, 1.5 h = beginning of zoospore germination, and 2.5 h = fully germinated zoospores. Alternative isoform expression was quantified using real-time PCR with transcript normalization to the *P. cubensis* internal transcribed spacer region. Real-time PCR is shown as the average with standard error of three biological replicates; dpi = days postinoculation and Sp = sporangia.

This hypothesis was developed based on a recent publication showing that transcripts containing introns were spliced after certain environmental cues, which allowed for rapid development to occur posttranscriptionally (Boothby et al. 2013). Therefore, we examined the changes for PCU\_115490 at three different points during the sporangia-to-zoospore transition period that occurs under exposure to water, which triggers germination (Fig. 5). In sporangia (0 h) and during the early stages of zoospore formation (1.5 h), both isoforms were expressed at similar levels. At the latest stage of zoospore formation, PCU\_115490.1 accumulated to higher levels than PCU\_115490.2.

Previously, we reported on the role of AS in generating a candidate effector protein (Savory et al. 2012b). Our bioinformatics approach described in this study successfully identified PscRXLR1 (PCU\_071030) as a candidate RXLR with evidence for AS. We further suggest that nearly 1% of the alternatively spliced genes are predicted to encode effector proteins. Within this class of genes, only intron retention was identified as an AS mechanism. PCU\_240880 encodes a predicted RXLR effector. The retained intron inferred from RNA-Seq occurs in its 3' UTR and is not predicted to affect function but it may regulate expression (Fig. 6A). Both of the gene isoforms show similar patterns of expression, with peak levels observed at 2 dpi (Fig. 6A). More importantly, the ratios of PCU\_240880.1 to PCU\_240880.2 changed, because more PCU\_240880.1 could be detected at higher levels in *P. cubensis* during its host-associated stages (>60%) compared with sporangia (approximately 40%). PCU\_143920 is also predicted to encode an RXLR effector. This gene has two introns and data suggest that only the second intron is retained (Fig. 6B). The relative expression of PCU\_143920.1 significantly decreased ( $t$  test  $P$  value < 0.05) in levels from sporangia to the host-associated stages (Fig. 6B). Likewise, the relative expression of PCU\_143920.2 also decreased but its overall relative expression levels were consistently low. Finally, PCU\_031660 is predicted to encode a QXLR effector. It has two introns, and inferences from RNA-Seq data suggest that the first intron is retained, with the strongest evidence being from sporangia tissue (Fig. 6C). The relative expression of PCU\_031660.1 significantly increased ( $t$  test  $P$  value < 0.05) from sporangia to the host-associated stages of *Pseudoperonospora cubensis*. In contrast, the expression of PCU\_031660.2 remained relatively low and constant.

## DISCUSSION

In this study, we used RNA-Seq to characterize the transcriptome of *P. cubensis*. We sampled multiple time points during the infection cycle of *P. cubensis* on cucumber. This enabled us to capture gene expression in multiple pathogen structures, including sporangia, encysted zoospores, appressoria, hyphae, and haustoria (Savory et al. 2012a). By using the Illumina HiSeq, we were able to obtain approximately 40 million sequenced fragments that uniquely aligned to *P. cubensis*, more than doubling the number of usable sequences that were previously generated using the Illumina GAI (Savory et al. 2012a). Inferences based on unique alignments revealed approximately 17,500 expressed genes, including approximately 4,000 new genes. This enabled us to both improve the existing genome annotation and provide sufficient coverage for the first transcriptome-wide view into AS of an obligate oomycete plant pathogen.

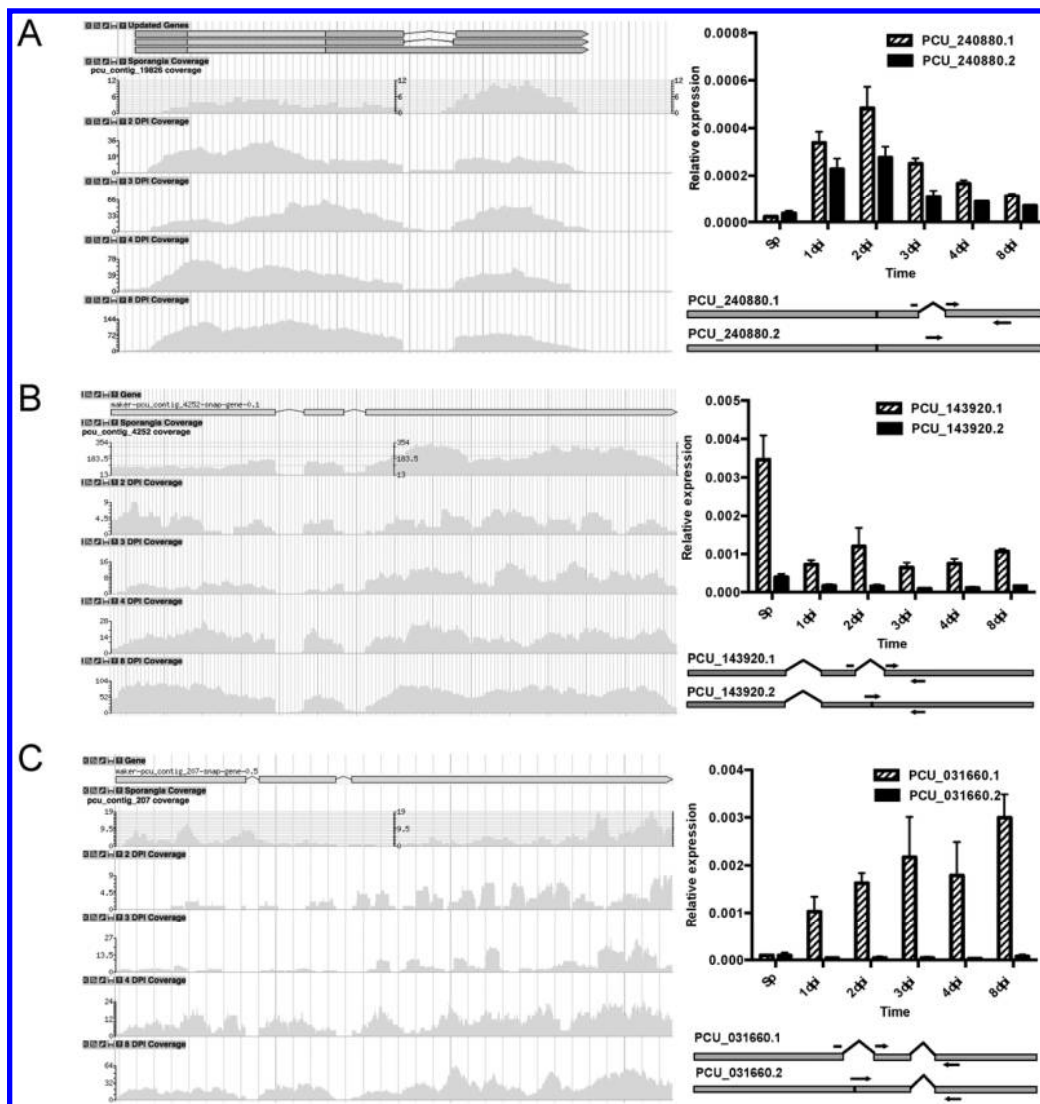
The original draft sequence of *P. cubensis* consists of a large number of contigs, 23,519 annotated genes, and few predicted gene models. With the RNA-Seq reads, we could support 57% of those genes, and also identified over 4,000 previously unannotated genes. Moreover, more than half of the previously annotated genes were modified to include additional coding sequences, UTR, or modifications to their gene models.



The current draft genome sequence, improved using RNA-Seq datasets, has 27,591 annotated genes, of which 64% are predicted to be expressed. Although this may seem low, it is likely due to the fact that not all stages of the *P. cubensis* lifestyle, such as sporulation and sexual reproduction, were represented in the sampled time points. An additional confounding factor is the relatively low depth of transcriptome sequencing caused by the high percentage of host transcripts in each sample, which again highlights one of the challenges of studying this pathosystem. Additionally, or alternatively, the low percentage of expressed genes could also suggest that the original genome sequence included contaminating sequences and that many of the annotated genes were erroneously predicted. Nonetheless, our analysis indicates that the original draft genome of *P. cubensis* did not likely suffer from missing gene sequences. We were able to align nearly 10 million sequenced fragments that were previously unannotated to the reference sequence. Most of the unaligned read sequences were determined to be the result of sequencing errors and not predicted to be from *P. cubensis*.

Thus, the use of RNA-Seq led to dramatic changes to the reference sequence and contributed to a higher quality resource for advancing our understanding of this critically important pathogen.

The extent and contribution of AS to transcriptome and proteome diversity in Stramenopiles is less characterized than in fungi, plants, and mammals. Results from this study suggested that AS occurs with a much higher frequency than had been originally predicted. In *P. cubensis*, we identified 24% of the expressed genes as having evidence for AS (Table 1). These data are not consistent with previous work that predicted lower rates of AS in plant-pathogenic oomycetes (McGuire et al. 2008; Shen et al. 2011). However, a critical difference between the current work and previously published studies is the technology used in transcriptome sampling. The current work, which relied on RNA-Seq, provided a significantly deeper and more comprehensive data set from which the *P. cubensis* transcriptome can be modeled for AS, in spite of the overwhelming number of reads that were derived from the cucumber host.



**Fig. 6.** Predicted alternatively spliced genes encoding secreted gene products predicted to contain a signal peptide and an RXLR or QXLR motif. RNA-Seq coverage of each gene is shown as visualized through GBrowse. Coverage is shown for sporangia, 2, 3, 4, and 8 days postinoculation (dpi). Gene models are provided for each isoform, with a vertical bar indicating the location of a stop codon. Real-time polymerase chain reaction was used to determine the relative expression of alternatively spliced isoforms as normalized to the *Pseudoperonospora cubensis* internal transcribed spacer region. Arrows on the models indicate the binding position of isoform-specific primers. Data are reported as the average relative expression and standard error from three biological replicates. **A**, PCU\_240880 encodes a predicted RXLR effector and is alternative spliced via the mechanism of intron retention in the 3' untranslated region. **B**, PCU\_143920 encodes a predicted RXLR effector and is alternative spliced in the coding sequence (CDS) of the gene. **C**, PCU\_031660 encodes a predicted QXLR effector whose CDS is alternatively spliced.

Indeed, RNA-Seq-based analysis of *Bigeloviella natans*, a chlorarachniophyte alga, revealed that AS occurs at levels remarkably higher than previously observed for other unicellular eukaryotes (Curtis et al. 2012). This observation is significant because oomycetes are more closely related to *B. natans* and plants than to fungi (Burki et al. 2012; Hackett et al. 2007; Seidel et al. 2011; Simpson and Roger 2004). Thus, it is distinctly possible that AS in oomycetes occurs at rates more in line with occurrence in plants.

The higher frequency of intron retention observed in *P. cubensis* is consistent with previous studies of AS in oomycetes (McGuire et al. 2008; Shen et al. 2011), and is consistent with high intron retention rates identified in plants (Filichkin et al. 2010). Intron retention has been proposed as a mechanism of autoregulation for controlling both the timing and expression of the genome. The splicing of introns can be regulated in a stage- and stress-dependent manner, as has been demonstrated for influencing the expression of meiotic genes and ribosomal protein-encoding genes in response to amino acid starvation (Juneau et al. 2007; Pleiss et al. 2007; Wilhelm et al. 2008). In the fern *Marsilea vestita*, intron retention has been shown to be important for regulating gamete development, whereby transcripts with retained introns prevent the early translation of a protein, but upon developmental cues, are rapidly spliced to trigger developmental changes (Boothby et al. 2013). In *P. cubensis*, notable changes in splicing patterns occurred during the transition from sporangia to host-associated life stages. This is consistent with a model suggesting that AS can suppress translation and maintain the oomycete in a quiescent but poised state that, upon splicing, contributes to a rapid transition into an infective stage (Figs. 4 to 6).

A proportion of the alternative spliced genes could lead to different protein isoforms. In *Phytophthora sojae*, alternatively spliced Crinkler effector transcripts have been reported (Shen et al. 2011). Similarly, an effector gene from *Glomus intraradices* is alternatively spliced to generate multiple isoforms that vary in the number of C-terminal repeats (Kloppholz et al. 2011). In *Colletotrichum lindemuthianum* and *Botrytis cinerea*, exon skipping results in the exclusion of a zinc finger domain from the STE12 transcription factor (Hoi et al. 2007; Schamber et al. 2010). As a function of the host–pathogen interaction in the case of *C. lindemuthianum*, expression of the alternatively spliced truncated form of STE12 has been shown to negatively affect pathogenicity through a downregulation in expression coincident with appressorium formation (Hoi et al. 2007). In future studies, it will be important to identify and characterize genes that give rise to different protein isoforms that contribute to the virulence of *Pseudoperonospora cubensis*.

The relevance of the extraordinarily high numbers of estimated alternatively spliced transcripts has been repeatedly challenged. At the core of the debate is the fate of the alternatively spliced transcript. Many isoforms are predicted to encode non-functional proteins and, regardless of functionality, most lack support from expressed protein variants in databases (Hegyi et al. 2011). As a consequence, sloppiness in splicing (i.e., stochastic noise) is often invoked to explain AS and the generation of large numbers of seemingly nonproductive transcripts (Melamud and Moulton 2009; Pickrell et al. 2010). The contributions of stochastic noise and technical biases in estimates of AS in the RNA-Seq datasets described herein are difficult to quantify. In humans, for example, splicing errors are conservatively estimated to occur in approximately 2% of the transcripts from the average genes (Pickrell et al. 2010). However, this estimate was made using unannotated spliced sequences that were poorly supported, represented by only 1.7% of the gap-aligned RNA-Seq read sequences. In this study, a number of filters were implemented to limit poorly supported AS events and reduce the

misclassification of noise as relevant AS events. As a result of our stringent analysis, the alternative 5' and 3' splice-type events were supported by 6.3% of the gap-aligned RNA-Seq read sequences. The prediction of intron retention was more challenging and could be a consequence of sequencing pre-mRNAs. However, in PCU\_071910 and PCU\_143920, retention was only observed for one of their two introns, arguing against artifacts. Conservatively, because AS is often developmentally regulated or stress induced, we hypothesize that the number of splice events in the *P. cubensis* transcriptome and in the transcriptomes of other oomycetes is actually higher than predicted. To fully define the role of AS in the regulation of pathogen development, in response to environment, and in mediation of host defense signaling, future studies are needed to examine the splicing variation under differing conditions; these include abiotic stresses (e.g., changes in temperature) and biotic stresses such as variation of host range and differing levels of host resistance. As presented herein, the current work provides a foundation for such studies, which will help reveal the true extent of AS and better understand the role of AS in oomycete development and virulence of plant hosts.

## MATERIALS AND METHODS

### Plant growth and *P. cubensis* inoculation.

Uninfected 'Vlaspik' cucumber plants were grown at 22°C in cycles of 16 h of light and 8 h of darkness. *P. cubensis* isolate MSU-1 was propagated on cucumber, as previously described (Tian et al. 2011). Time course experiments were performed as described by Savory and associates (2012a), in which 10 µl of a suspension of *P. cubensis* at  $1 \times 10^5$  sporangia/ml in distilled water was drop inoculated onto the abaxial leaf surface of 4- to 5-week-old cucumber plants. Infected plants were kept in the dark at 100% humidity for 24 h, and were subsequently maintained at near 100% humidity at 22°C in cycles of 12 of light and 12 h of darkness (Savory et al. 2012a). Infected leaf samples were collected at the same time each day of the time course using a number 3 cork borer, flash-frozen in liquid nitrogen, and stored at –80°C until sample processing.

### Next-generation sequencing.

Previously prepared samples (Savory et al. 2012a) were resequenced more deeply for this study. The libraries from biological replicates of cucumber leaves infected with *P. cubensis* for 2, 3, 4, and 8 dpi were resequenced using 100-mer PE sequencing on four channels of an Illumina HiSeq at the Michigan State University Research Technology Support Facility. Previously sequenced *P. cubensis* sporangia sample reads (Savory et al. 2012a) were used for bioinformatic analyses. The short reads are linked to Bioproject PRJNA247812 and also available from the Short Read Archive available through the National Center for Biotechnology Information (SRP042019).

### Alignment and processing of RNA-Seq read sequences.

Bowtie version 0.12.7 was used to align RNA-Seq read sequences with up to two mismatches to the *Cucumis sativus* L. (genome and contig accessions ACHR000000000 to ACHR01059995) and *P. cubensis* (genome and contig accessions AHJF000000000 to AHJF01035539) reference sequences (Huang et al. 2009; Savory et al. 2012a). PE RNA-Seq reads were aligned independently. The quality of the reads was determined by assessing mismatches as a factor of nucleotide position. Based on the bias of mismatches toward the ends of the sequencing reads, between 5 and 31 nt were trimmed from the ends. Using a custom C++ pipeline, high-quality pairs were first identified based on the criteria that one read must align per strand and the alignments must be between –60 and

200 nt apart. A second step using an iterative process that adjusted the distance based on sizes of intervening annotated introns was employed to identify pairs that exceeded the distance criterion because of an annotated intron. SuperSplat modified to output SAM-formatted data was used for gapped alignments (Bryant et al. 2010). Alignment parameters included a perfect match on either side of the gap, with a minimum of at least 15 nt on one side of the gap and a gap of 20 to 4,000 nt in length. All alignments were stored in BAM format using SAMtools (Li et al. 2009). Both PE and a subset of single, unpaired RNA-Seq reads were used, which are collectively referred to as “sequenced fragments”.

### Gene improvement and discovery.

Custom Perl and Python scripts were developed to update and identify genes based on coverage and predicted splice junction sequences from the aggregated sequenced fragments that uniquely aligned to the *P. cubensis* reference genome sequence. Genes in the original annotation were extended if adjacent positions had  $\geq 10\times$  coverage or had a gap-aligned fragment sequence that was consistent with it being on the sense strand of the gene (Supplementary Fig. S5). New genes were discovered by identifying regions that did not overlap features in the original annotation and met the criteria of  $\geq 300$  nt in length, a coverage  $\geq 10\times$ , and no gaps in coverage  $>5$  nt that were not supported by gap alignments. For loci that contained gap-aligned fragment sequences, the dinucleotide sequence was used to predict the strand of the gene. Open reading frames were scanned in all three frames and the largest of the potential coding sequences that exceeded 90 amino acids were annotated using IPRScan and BLAST (Altschul et al. 1990; Zdobnov and Apweiler 2001). A final script was used to merge results, remove duplicates, and format the final output using the Generic Feature Format Version 3 format specification. The updated sequence annotation is available for download from Oregon State University Libraries.

### Comparative analysis of AS among eukaryotes.

Custom Perl scripts were used to calculate the percentage of single exonic and intron-bearing genes, ratios of AS types, and percentage of total AS from representative eukaryotic genomes. Genome annotation information was obtained from the appropriate genome databases.

### Prediction of candidate secreted and effector proteins.

SignalP (version 4.0; using the noTM network with a D-cut-off of 0.34) was used to predict signal peptide cleavage sites in translated sequences to identify candidate secreted proteins (Petersen et al. 2011). A string search for “RXLR” within a distance of 90 amino acids of a predicted signal peptide cleavage site was used to query the candidate-secreted proteins for putative effector proteins.

### Prediction of alternatively spliced isoforms.

To identify potential intron retention events, only introns in which the coverage of the flanking 30 nt  $\geq 10\times$  were considered as a binary measure of expression. The “local intron coverage” ratios were calculated by normalizing the coverage of introns to the coverage of the flanking 30 nt. Retained introns were expected to exceed a threshold of 13%, which was set based on it being nearly two standard deviations to the right of the mean local intron coverage of all expressed introns. We also required intron retention events to be supported by reads that spanned the spliced intron as well as those that aligned to the exon-intron border sequence. To distinguish an alternatively spliced sequence from an erroneous annotation, we required additional correspondence of putatively spliced sequences to an intronic

region that had support for at least one other spliced variant affecting the same region (e.g., the annotated splice junction and an alternatively spliced sequence). Finally, the local intron coverage ratio could not exceed 80% to minimize the categorization of misannotated exonic sequences as intron retention events. All of the predicted intron retention events were confirmed using an independent approach. Intron retention events were inferred based on the unambiguous mapping of reads to exon-intron junctions flanking an intron and the gapped mapping of reads to exon-exon junctions that spanned the intron.

### RNA and DNA isolation.

The RNeasy plant mini kit (Qiagen, Germantown, MD, U.S.A.) was used to extract total RNA from sporangia and flash-frozen leaf tissue. An on-column DNase treatment (Qiagen) was used to remove contaminating genomic DNA from the RNA samples. The Qiagen RNeasy plant mini kit was used to extract genomic DNA from sporangia.

### RT-PCR analysis.

First-strand cDNA was synthesized from RNA isolated from sporangia and infected plant tissue using oligo-dT primers and the first-strand cDNA synthesis kit (USB Affymetrix, Santa Clara, CA, U.S.A.). RT-PCR was performed using GoTaq (Promega Corp., Madison, WI, U.S.A.). DNA primer sequences used for amplification are listed in Supplementary Table S2. The following cycling parameters were used: 1 hold at 95°C (1 min); 50 cycles of 95°C (30 s), 55°C (1 min), and 72°C (3 min); and 1 hold at 72°C (10 min). RT-PCR products were separated on a 1.5% agarose gel with an Invitrogen 1-kb Plus DNA ladder and stained with ethidium bromide. Genomic *P. cubensis* DNA was used as a positive PCR control. PCR products were sequenced using the ABI 3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, U.S.A.) at the Michigan State University Research Technology Support Facility to confirm that the sequence matched the predicted spliced sequence.

### Quantitative real-time PCR.

First-strand cDNA was synthesized using the USB first-strand cDNA synthesis kit with random hexamer primers (USB Affymetrix), as described above, using 1  $\mu$ g of total RNA. Samples were prepared using the HotStart SYBR Green qPCR Master Mix (2 $\times$ ; USB) and the DNA primers listed in Supplementary Table S3. Quantitative real-time PCR was performed using a Mastercycler ep Realplex real-time PCR machine (Eppendorf AG, Hamburg, Germany). *P. cubensis*-specific internal transcribed spacer (ITS) was also amplified (Tian et al. 2011) and used to normalize for the level of *P. cubensis* in infected tissue samples. The following cycling parameters were used: 1 hold of 95°C for 2 min and 40 cycles of 95°C (15 s), 56°C (15 s), and 72°C (30 s). Relative expression for each splice variant was calculated where Ct is the cycle threshold value and relative expression =  $2^{-(\Delta Ct)}$  and where  $\Delta Ct = Ct_{\text{gene of interest}} - Ct_{\text{ITS}}$ , as previously described (Porter et al. 2012). A melt-curve analysis was run for each primer set to ensure that a unique product was amplified. Error bars represent the standard error from three biological replicates. Data were analyzed and processed using Prism (GraphPad Software, Inc., San Diego, CA, U.S.A.). Unpaired two-tailed *t*-test analyses were performed using the mean and standard error of the mean for pairwise sample comparisons with a significance *P* value cutoff of 0.05.

### Transient gene expression and protein localization in *N. benthamiana*.

PCR products of the constitutively and alternatively spliced transcripts (to the first premature stop codons) of PCU\_071910

and PCU\_115490 were amplified in RT-PCRs and ligated to the pENTR/D-TOPO Gateway entry vector (Invitrogen, Carlsbad, CA, U.S.A.). The cDNAs were recombined into pVKH18-En6gw-cCfP (Tian et al. 2011) using LR Clonase (Invitrogen). DNA sequences were confirmed using the ABI 3730 Genetic Analyzer (Applied Biosystems).

Plasmid constructs were electroporated into *Agrobacterium tumefaciens* C58C1. For in planta localization, 4- to 6-week-old *N. benthamiana* were co-infiltrated with strains of *Agrobacterium* carrying pVKH18En6gw-cCfP variants paired with a strain carrying *AtPIP2A-YFP* (Nelson et al. 2007; Savory et al. 2012b). The latter encodes a translational fusion between a plasma membrane marker and yellow fluorescent protein (YFP). Images were collected 48 h postinfiltration using the Olympus FV1000 laser-scanning confocal microscope (Olympus America, Center Valley, PA, U.S.A.). Images for negative controls were also collected to ensure that there was no fluorescence crossover; cells transiently expressing *AtPIP2A-YFP* were not detectable when visualized using the CFP channel settings used for image collection.

### Sporangia time course sampling.

Ungerminated sporangia emerging from sporangiophores on leaf surfaces did not contact water, and thus did not germinate prior to flash-freezing; these samples were considered 0 h sporangia. Flash-frozen leaf tissues harboring 0 h sporangia were shaken in flasks of distilled water at 120 rpm at room temperature to dislodge sporangia. Samples were filtered through a 70-micron filter to remove debris, and centrifuged at  $3,220 \times g$  to pellet the sporangia. Sporangia samples from multiple tubes were pooled and centrifuged again at 4,000 rpm in a microcentrifuge to further pellet the sporangia for RNA extraction. Isolated samples were visually inspected using light microscopy to ensure that flash-frozen sporangia did not germinate during the sample preparation process. To collect sporangia for zoospore germination, fresh sporulating leaves were processed as described above, except that infected leaf tissues were not flash-frozen. Biological samples were collected after 1.5 and 2.5 h in sterile distilled water. Germination of sporangia was observed using an Olympus IX71 inverted light microscope. After 1.5 h in water, 30 to 50% of the sporangia were germinating into zoospores and, after 2.5 h post-water exposure, nearly 100% of the sporangia had germinated. Three biological replicates of this time course were collected.

### ACKNOWLEDGMENTS

We thank S. Cusack for assistance with sample preparation. E. A. Savory is supported by United States Department of Agriculture NIFA postdoctoral fellowship number 2013-67012-21139. Work in the laboratory of J. H. Chang is supported by funding from the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R01GM104977. Research in the laboratory of B. Day is supported by funding from the Michigan State University Rackham Foundation, Michigan State University Project GREEN (GR13-007), and a grant from the United States Department of Agriculture Specialty Crops Research Initiative (2011-51181-30661).

### LITERATURE CITED

- Adhikari, B. N., Savory, E. A., Vaillancourt, B., Childs, K. L., Hamilton, J. P., Day, B., and Buell, C. R. 2012. Expression profiling of *Cucumis sativus* in response to infection by *Pseudoperonospora cubensis* PLoS One 7:e34954. Published online.
- ▶ Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- ▶ Beakes, G. W., Glockling, S. L., and Sekimoto, S. 2012. The evolutionary phylogeny of the oomycete “fungi”. *Protoplasma* 249:3-19.
- ▶ Boothby, T. C., Zipper, R. S., van der Weele, C. M., and Wolniak, S. M. 2013. Removal of retained introns regulates translation in the rapidly developing gametophyte of *Marsilea vestita*. *Dev. Cell* 24:517-529.
- ▶ Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otiillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J.-P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kröger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Jézéquel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M.-P., Napoli, C., Obornik, M., Parker, M. S., Petit, J.-L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siat, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardil, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S., Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van de Peer, Y., and Grigoriev, I. V. 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456:239-244.
- ▶ Bryant, D. W., Shen, R., Priest, H. D., Wong, W. K., and Mockler, T. C. 2010. Supersplat—spliced RNA-seq alignment. *Bioinformatics* 26:1500-1505.
- ▶ Burkhardt, A., and Day, B. 2013. A genomics perspective on cucurbit-oomycete interactions. *Plant Biotechnol.* 30:265-271.
- ▶ Burki, F., Okamoto, N., Pombert, J.-F., and Keeling, P. J. 2012. The evolutionary history of haptophytes and cryptophytes: Phylogenomic evidence for separate origins. *Proc. R. Soc. B Biol. Sci.* 279:2246-2254.
- ▶ Chen, L., Tovar-Corona, J. M., and Urrutia, A.O. 2012. Alternative splicing: A potential source of functional innovation in the eukaryotic genome. *Int. J. Evol. Biol.* 2012:1110.
- ▶ Curtis, B. A., Tanifuji, G., Burki, F., Gruber, A., Irimia, M., Maruyama, S., Arias, M. C., Ball, S. G., Gile, G. H., Hirakawa, Y., Hopkins, J. F., Rensing, S. A., Schmutz, J., Symeonidi, A., Elias, M., Eveleigh, R. J., Herman, E. K., Klute, M. J., Nakayama, T., Obornik, M., Reyes-Prieto, A., Armbrust, E. V., Aves, S. J., Beiko, R. G., Coutinho, P., Dacks, J. B., Durnford, D. G., Fast, N. M., Green, B. R., Grisdale, C. J., Hempel, F., Henrissat, B., Höppner, M. P., Ishida, K., Kim, E., Kořený, L., Kroth, P. G., Liu, Y., Malik, S. B., Maier, U. G., McRose, D., Mock, T., Neilson, J. A., Onodera, N. T., Poole, A. M., Pritham, E. J., Richards, T. A., Rocap, G., Roy, S. W., Sarai, C., Schaack, S., Shirato, S., Slamovits, C. H., Spencer, D. F., Suzuki, S., Worden, A. Z., Zauner, S., Barry, K., Bell, C., Bharti, A. K., Crow, J. A., Grimwood, J., Kramer, R., Lindquist E., Lucas, S., Salamov, A., McFadden, G. I., Lane, C. E., Keeling, P. J., Gray, M. W., Grigoriev, I. V., and Archibald J.M. 2012. Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. *Nature* 492:59-65.
- ▶ Filichkin, S. A., Priest, H. D., Givan, S. A., Shen, R., Bryant, D. W., Fox, S. E., Wong, W. K., and Mockler, T. C. 2010. Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. *Genome Res.* 20:45-58.
- ▶ Guan, Q., Wu, J., Zhang, Y., Jiang, C., Liu, R., Chai, C., and Zhu, J. 2013. A DEAD box RNA helicase is critical for pre-mRNA splicing, cold-responsive gene regulation, and cold tolerance in *Arabidopsis*. *Plant Cell* 25:342-356.
- ▶ Hackett, J. D., Yoon, H. S., Li, S., Reyes-Prieto, A., Rümmele, S. E., and Bhattacharya, D. 2007. Phylogenomic analysis supports the monophyly of cryptophytes and haptophytes and the association of rhizaria with chromalveolates. *Mol. Biol. Evol.* 24:1702-1713.
- ▶ Hegyi, H., Kalmár, L., Horváth, T., and Tompa, P. 2011. Verification of alternative splicing variants based on domain integrity, truncation length and intrinsic protein disorder. *Nucleic Acids Res.* 39:1208-1219.
- ▶ Hoi, J. W. S., Herbert, C., Bacha, N., O’Connell, R., Lafitte, C., Borderies, G., Rossignol, M., Rougé, P., and Dumas, B. 2007. Regulation and role of a STE12-like transcription factor from the plant pathogen *Colletotrichum lindemuthianum*. *Mol. Microbiol.* 64:68-82.
- Holmes, G., Wehner, T., and Thornton, A. 2006. An old enemy re-emerges. *Am. Veg. Grow.* February:14-15.
- ▶ Hoskins, A. A., and Moore, M. J. 2012. The spliceosome: A flexible, reversible macromolecular machine. *Trends Biochem. Sci.* 37:179-188.
- ▶ Huang, S., Li, R., Zhang, Z., Li, L., Gu, X., Fan, W., Lucas, W.J., Wang, X., Xie, B., Ni, P., Ren, Y., Zhu, H., Li, J., Lin, K., Jin, W., Fei, Z., Li, G., Staub, J., Kilian, A., van der Vossen, E., Wu, y., Gui, J., He, J., Jia, Z., Ren, Y., Tian, G., Lu, Y., Ruan, J., Qian, E., Wang, M., Huang, Q., Li, B., Xua, Z., Cao, J., Asan, Wu, Z., Cai, Q., Bai, Y., Zhao, B., Han, Y., Li, Y., Li, X., Wang, S., Shi, Q., Liu, S., Cho, W. K., Kim, J. Y., Xu, Y., Heller-Uzyska, K., Miao, H., Cheng, Z., Zhang, S., Wu, J., Yang, Y., Kang, H., Li, M., Liang, H., Ren, X., Shi, Z., Wen, M., Jian, M., Yang, H., Zhang, G., Yang, Z., Chen, R., Liu, S., Li, J., Ma, L., Liu, S., Li, J., Zhou, Y., Zhao, J., Fang, X., Li, G., Fang, L., Li, Y., Liu, D., Zheng, H., Zhang, Y., Qin, N., Li, Z., Yang, G., Yang, S., Bolund, L., Kristiansen, K., Zheng, H., Li, S., Zhang, X., Yang, H., Want, J., Sun, R., Zhang, B., Jiang, S., Wang,



- J., Du, Y., and Li, S. 2009. The genome of the cucumber, *Cucumis sativus* L. *Nat. Genet.* 41:1275-1281.
- ▶ Irimia, M., Penny, D., and Roy, S. W. 2007. Coevolution of genomic intron number and splice sites. *Trends Genet.* 23:321-325.
  - ▶ James, A. B., Syed, N. H., Bordage, S., Marshall, J., Nimmo, G. A., Jenkins, G. I., Herzyk, P., Brown, J. W. S., and Nimmo, H. G. 2012. Alternative splicing mediates responses of the *Arabidopsis* circadian clock to temperature changes. *Plant Cell* 24:961-981.
  - ▶ Juneau, K., Palm, C., Miranda, M., and Davis, R. W. 2007. High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc. Natl. Acad. Sci. U.S.A.* 104:1522-1527.
  - ▶ Kalsotra, A., and Cooper, T. A. 2011. Functional consequences of developmentally regulated alternative splicing. *Nat. Rev. Genet.* 12:715-729.
  - ▶ Kamoun, S. 2006. A catalogue of the effector secretome of plant pathogenic oomycetes. *Annu. Rev. Phytopathol.* 44:41-60.
  - ▶ Keren, H., Lev-Maor, G., and Ast, G. 2010. Alternative splicing and evolution: Diversification, exon definition and function. *Nat. Rev. Genet.* 11:345-355.
  - ▶ Kim, E., Goren, A., and Ast, G. 2008. Alternative splicing: Current perspectives. *BioEssays* 30:38-47.
  - ▶ Klopffholz, S., Kuhn, H., and Requena, N. 2011. A secreted fungal effector of *Glomus intraradices* promotes symbiotic biotrophy. *Curr. Biol.* 21:1204-1209.
  - ▶ Labadorf, A., Link, A., Thomas, J., Reddy, A. S., and Ben-Hur, A. 2010. Genome-wide analysis of alternative splicing in *Chlamydomonas reinhardtii*. *BMC Genomics* 11:114.
  - ▶ Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078-2079.
  - ▶ Marquez, Y., Brown, J. W. S., Simpson, C., Barta, A., and Kalyna, M. 2012. Transcriptome survey reveals increased complexity of the alternative splicing landscape in *Arabidopsis*. *Genome Res.* 22:1184-1195.
  - ▶ Mastrangelo, A. M., Marone, D., Laido, G., De Leonardi, A. M., and De Vita, P. 2012. Alternative splicing: Enhancing ability to cope with stress via transcriptome plasticity. *Plant Sci.* 185-186:40-49.
  - ▶ McGuire, A. M., Pearson, M. D., Neafsey, D. E., and Galagan, J. E. 2008. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biol.* 9:R50.
  - ▶ Melamud, E., and Moul, J. 2009. Stochastic noise in splicing machinery. *Nucleic Acids Res.* 37:4873-4886.
  - ▶ Meyer, M., and Vilardell, J. 2008. The quest for a message: Budding yeast, a model organism to study the control of pre-mRNA splicing. *Brief. Funct. Genomics Proteomics* 8:60-67.
  - ▶ Nelson, B. K., Cai, X., and Nebenführ, A. 2007. A multicolored set of in vivo organelle markers for co-localization studies in *Arabidopsis* and other plants. *Plant J.* 51:1126-1136.
  - ▶ Nicholson, P., Yepiskoposyan, H., Metz, S., Orozco R. Z., Kleinschmidt, N., and Mühlemann, O. 2009. Nonsense-mediated mRNA decay in human cells: Mechanistic insights, functions beyond quality control and the double-life of NMD factors. *Cell Mol. Life Sci.* 67:677-700.
  - ▶ Pan, Q., Shai, O., Lee, L. J., Frey, B. J., and Blencowe, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* 40:1413-1415.
  - ▶ Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. 2011. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* 8:785-786.
  - ▶ Pickrell, J. K., Pai, A. A., Gilad, Y., and Pritchard, J. K. 2010. Noisy splicing drives mRNA isoform diversity in human cells. *PLoS Genet.* 6:e1001236. Published online.
  - ▶ Pistoni, M., Ghigna, C., and Gabellini, D. 2010. Alternative splicing and muscular dystrophy. *RNA Biol.* 7:441-452.
  - ▶ Pleiss, J. A., Whitworth, G. B., Bergkessel, M., and Guthrie, C. 2007. Rapid, transcript-specific changes in splicing in response to environmental stress. *Mol. Cell* 27:928-937.
  - ▶ Porter, K., Shimono, M., Tian, M., and Day, B. 2012. *Arabidopsis* actin-depolymerizing factor-4 links pathogen perception, defense activation and transcription to cytoskeletal dynamics. *PLoS Pathog.* 8:e1003006. Published online.
  - ▶ Rayson, S., Arciga-Reyes, L., Wootton, L., De Torres-Zabala, M., Truman, W., Graham, N., Grant, M., and Davies, B. 2012. A role for nonsense-mediated mRNA decay in plants: Pathogen responses are induced in *Arabidopsis thaliana* NMD mutants. *PLoS One* 7:e31917. Published online.
  - ▶ Reddy, A. S. N. 2007. Alternative splicing of pre-messenger RNAs in plants in the genomic era. *Annu. Rev. Plant Biol.* 58:267-294.
  - ▶ Rehmany, A., Gordon, A., Rose, K. E., Allen, R. L., Armstrong, M. R., Whisson, S. C., Kamoun, S., Tyler, B. M., Birch, P. R. J., and Beynon, J. L. 2005. Differential recognition of highly divergent downy mildew avirulence gene alleles by *RPP1* resistance genes from two *Arabidopsis* lines. *Plant Cell* 17:1839-1850.
  - ▶ Sanchez, S. E., Petrillo, E., Kornblihtt, A. R., and Yanovsky, M. J. 2011. Alternative splicing at the right time. *RNA Biol.* 8:954-959.
  - ▶ Savory, E. A., Granke, L. L., Quesada-Ocampo, L. M., Varbanova, M., Hausbeck, M. K., and Day, B. 2010. The cucurbit downy mildew pathogen *Pseudoperonospora cubensis*. *Mol. Plant Pathol.* 12:217-226.
  - ▶ Savory, E. A., Adhikari, B. N., Hamilton, J. P., Vaillancourt, B., Buell, C. R., and Day, B. 2012a. mRNA-Seq analysis of the *Pseudoperonospora cubensis* transcriptome during cucumber (*Cucumis sativus* L.) infection. *PLoS One* 7:e35796. Published online.
  - ▶ Savory, E. A., Zou, C., Adhikari, B. N., Hamilton, J. P., Buell, C. R., Shiu, S.-H., and Day, B. 2012b. Alternative splicing of a multi-drug transporter from *Pseudoperonospora cubensis* generates an RXLR effector protein that elicits a rapid cell death. *PLoS One* 7:e34701. Published online.
  - ▶ Schamber, A., Leroch, M., Diwo, J., Mendgen, K., and Hahn, M. 2010. The role of mitogen-activated protein (MAP) kinase signalling components and the Ste12 transcription factor in germination and pathogenicity of *Botrytis cinerea*. *Mol. Plant Pathol.* 11:105-119.
  - ▶ Schöning, J. C., Streitner, C., Page, D. R., Hennig, S., Uchida, K., Wolf, E., Furuya, M., and Staiger, D. 2007. Auto-regulation of the circadian slave oscillator component AtGRP7 and regulation of its targets is impaired by a single RNA recognition motif point mutation. *Plant J.* 52:1119-1130.
  - ▶ Schwartz, S. L., Cao, C., Pylypenko, O., Rak, A., and Wandinger-Ness, A. 2008. Rab GTPases at a glance. *J. Cell Sci.* 120:3905-3910.
  - ▶ Seidl, M. F., Van den Ackerveken, G., Govers, F., and Snel, B. 2011. A domain-centric analysis of oomycete plant pathogen genomes reveals unique protein organization. *Plant Physiol.* 155:628-644.
  - ▶ Shen, D., Ye, W., Dong, S., Wang, Y., and Dou, D. 2011. Characterization of intronic structures and alternative splicing in *Phytophthora sojae* by comparative analysis of expressed sequence tags and genomic sequences. *Can. J. Microbiol.* 57:84-90.
  - ▶ Simpson, A. G. B., and Roger, A. J. 2004. The real 'kingdoms' of eukaryotes. *Curr. Biol.* 14:R693-R696.
  - ▶ Sultan, M., Schulz, M. H., Richard, H., Magen, A., Klingenhoff, A., Scherf, M., Seifert, M., Borodina, T., Soldatov, A., Parkhomchuk, D., Schmidt, D., O'Keefe, S., Haas, S., Vingron, M., Lehrach, H., and Yaspo, M. L. 2008. A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321:956-960.
  - ▶ Syed, N. H., Kalyna, M., Marquez, Y., Barta, A., and Brown, J. W. S. 2012. Alternative splicing in plants—Coming of age. *Trends Plant Sci.* 17:616-623.
  - ▶ Tian, M., Win, J., Savory, E., Burkhardt, A., Held, M., Brandizzi, F., and Day, B. 2011. 454 genome sequencing of *Pseudoperonospora cubensis* reveals effector proteins with a QXLR translocation motif. *Mol. Plant-Microbe Interact.* 24:543-553.
  - ▶ Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., and Holt, R. A. 2001. The sequence of the human genome. *Science* 291:1304-1351.
  - ▶ Wang, B., Guo, G., Wang, C., Lin, Y., Wang, X., Zhao, M., Guo, Y., He, M., Zhang, Y., and Pan, L. 2010. Survey of the transcriptome of *Aspergillus oryzae* via massively parallel mRNA sequencing. *Nucleic Acids Res.* 38:5075-5087.
  - ▶ Wang, B. B., and Brendel, V. 2006. Genomewide comparative analysis of alternative splicing in plants. *Proc. Natl. Acad. Sci. U.S.A.* 103:7175-7180.
  - ▶ Whisson, S. C., Boevink, P. C., Moleleki, L., Avrova, A. O., Morales, J. G., Gilroy, E. M., Armstrong, M. R., Grouffaud, S., van West, P., Chapman, S., Hein, I., Toth, I. K., Pritchard, L., and Birch, P. R. 2007. A translocation signal for delivery of oomycete effector proteins into host plant cells. *Nature* 450:115-119.
  - ▶ Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., Penkett, C. J., Rogers, J., and Bähler, J. 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453:1239-1243.
  - ▶ Zdobnov, E. M., and Apweiler, R. 2001. InterProScan--an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847-848.