Proteogenomic analysis greatly expands the identification of proteins related to reproduction in the apogamous fern dryopteris affinis ssp. affinis

Grossmann, Jonas; Fernández, Helena; Chaubey, Pururawa M; Valdés, Ana E; Gagliardini, Valeria; Cañal, María J; Russo, Giancarlo; Grossniklaus, Ueli

Abstract: Performing proteomic studies on non-model organisms with little or no genomic information is still difficult. However, many specific processes and biochemical pathways occur only in species that are poorly characterized at the genomic level. For example, many plants can reproduce both sexually and asexually, the first one allowing the generation of new genotypes and the latter their fixation. Thus, both modes of reproduction are of great agronomic value. However, the molecular basis of asexual reproduction is not well understood in any plant. In ferns, it combines the production of unreduced spores (diplospory) and the formation of sporophytes from somatic cells (apogamy). To set the basis to study these processes, we performed transcriptomics by next-generation sequencing (NGS) and shotgun proteomics by tandem mass spectrometry in the apogamous fern D. affinis ssp. affinis. For protein identification we used the public viridiplantae database (VPDB) to identify orthologous proteins from other plant species and new transcriptomics data to generate a "species-specific transcriptome database" (SSTDB). In total 1,397 protein clusters with 5,865 unique peptide sequences were identified (13 decoy proteins out of 1,410, protFDR 0.93% on protein cluster level). We show that using the SSTDB for protein identification increases the number of identified peptides almost four times compared to using only the publically available VPDB. We identified homologs of proteins involved in reproduction of higher plants, including proteins with a potential role in apogamy. With the increasing availability of genomic data from non-model species, similar proteogenomics approaches will improve the sensitivity in protein identification for species only distantly related to models.

DOI: [https://doi.org/10.3389/fpls.2017.00336](https://doi.org/10.3389/fpls.2017.00336)

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: [https://doi.org/10.5167/uzh-144589](https://doi.org/10.5167/uzh-144589)
Published Version

Originally published at:
Grossmann, Jonas; Fernández, Helena; Chaubey, Pururawa M; Valdés, Ana E; Gagliardini, Valeria; Cañal, María J; Russo, Giancarlo; Grossniklaus, Ueli (2017). Proteogenomic analysis greatly expands the identification of proteins related to reproduction in the apogamous fern dryopteris affinis ssp. affinis. Frontiers in Plant Science, 8:336.
DOI: [https://doi.org/10.3389/fpls.2017.00336](https://doi.org/10.3389/fpls.2017.00336)
Proteogenomic Analysis Greatly Expands the Identification of Proteins Related to Reproduction in the Apogamous Fern *Dryopteris affinis* ssp. *affinis*

Jonas Grossmann1, Helena Fernández2*, Pururawa M. Chaubey3†, Ana E. Valdés4,5†, Valeria Gagliardini3, María J. Cañal2, Giancarlo Russo1 and Ueli Grossniklaus3*

1Functional Genomics Center Zurich, Zürich, Switzerland, 2Area of Plant Physiology, Department of Organisms and Systems Biology (BCS), Oviedo University, Oviedo, Spain, 3Institute of Plant and Microbial Biology, Zurich-Basel Plant Science Center, University of Zurich, Zürich, Switzerland, 4Physiological Botany, Uppsala BioCenter, Uppsala University, Uppsala, Sweden, 5Linnean Centre for Plant Biology, Uppsala, Sweden

Performing proteomic studies on non-model organisms with little or no genomic information is still difficult. However, many specific processes and biochemical pathways occur only in species that are poorly characterized at the genomic level. For example, many plants can reproduce both sexually and asexually, the first one allowing the generation of new genotypes and the latter their fixation. Thus, both modes of reproduction are of great agronomic value. However, the molecular basis of asexual reproduction is not well understood in any plant. In ferns, it combines the production of unreduced spores (diplospory) and the formation of sporophytes from somatic cells (apogamy). To set the basis to study these processes, we performed transcriptomics by next-generation sequencing (NGS) and shotgun proteomics by tandem mass spectrometry in the apogamous fern *D. affinis* ssp. *affinis*. For protein identification we used the public viridiplantae database (VPDB) to identify orthologous proteins from other plant species and new transcriptomics data to generate a “species-specific transcriptome database” (SSTDB). In total 1,397 protein clusters with 5,865 unique peptide sequences were identified (13 decoy proteins out of 1,410, protFDR 0.93% on protein cluster level). We show that using the SSTDB for protein identification increases the number of identified peptides almost four times compared to using only the publically available VPDB. We identified homologs of proteins involved in reproduction of higher plants, including proteins with a potential role in apogamy. With the increasing availability of genomic data from non-model species, similar proteogenomics approaches will improve the sensitivity in protein identification for species only distantly related to models.

**Keywords:** apogamy, apomixis, *Dryopteris affinis* ssp. *affinis*, fern, gametophyte, proteogenomics
INTRODUCTION

Most angiosperms reproduce sexually through seeds, but there are examples of asexual seed formation (apomixis), where seeds form without meiosis and fertilization (Figure 1). Apomictic plants produce clonal embryos by sporophytic or gametophytic apomixis (Nogler, 1984; Koltonow and Grossniklaus, 2003). In sporophytic apomixis, the embryo forms directly from the somatic diploid ovule tissue (nucellus or integument). In gametophytic apomixis, the multicellular embryo sac may originate from two different cellular lineages leading to a broad categorization of this developmental program into diplospory and apospory. In diplospory, the embryo sac originates from the megaspore mother cell, either directly by mitosis or after restitution during meiosis, while in apospory the embryo sac originates from nucellar cells. In both cases, the asexual embryo develops from the unreduced egg cell without fertilization (parthenogenesis). Because apomixis allows the fixation of complex genotypes, including that of highly productive F1 hybrids, many researchers have extolled the tremendous potential that apomixis holds for plant improvement (Spillane et al., 2004). In apogamy, somatic cells of the gametophyte are reprogrammed to start the sporophytic developmental program. Apogamy does not occur naturally in angiosperms but is frequent in ferns (Yang and Zhou, 1992; Okano et al., 2009). Apogamy may be obligate, when gametophytes produce non-functional gametes, facultative, or induced by exogenous factors (Fernández et al., 1996; Menéndez et al., 2006a; Cordle et al., 2007). In obligate apogamy, endomitosis prior to meiosis serves to maintain the sporophytic chromosome number throughout the life cycle (Manton, 1950; Sheffield et al., 1983).

Over the last decade, several studies focusing on apomixis in model species of angiosperms concluded that sexual and apomictic pathways share gene expression profiles and, thus, common molecular regulatory features, indicating that they are not distinct pathways (Grossniklaus et al., 2001; Tucker et al., 2003). However, how somatic cells, either of sporophytic or gametophytic (apogamy) origin, become embryogenic is unknown. Apogamy in ferns is easy to observe and the gametophyte of apogamous ferns can be useful for comparison with the gametophytic events in angiosperms (Cordle et al., 2010). Although ferns receive comparatively little attention and genome sequences of ferns are so far, unavailable, it is accepted that we need to extend our analyses to more phylogenetic branches (Barker and Wolf, 2010). To date, only few fern species have been used to study developmental processes (Whittier, 1971; Wen et al., 1999; Salmi et al., 2005, 2010; Kazmierczak, 2010; Lopez and Renzaglia, 2014; Valledor et al., 2014; de Vries et al., 2015).

Dryopteris affinis (Lowe) Fraser-Jenkins ssp. affinis (Western scaly male fern) is a diploid, apomictic fern, which originated from a cross between the sexual ancestor of the extant apomict D. wallachiana (Wallach’s wood fern) and the sexual D. oreades (mountain male fern; Fraser-Jenkins, 1986). The gametophyte of this species forms male but no female reproductive organs and, when cultured in vitro, reproduces by apogamy. Once the gametophyte becomes heart-shaped, a brown organization center develops near to the apical indentation that directly forms an apogamous embryo sporophyte (Fernández et al., 1996; Menéndez et al., 2006a).

An alternative for examining gene expression in species without a genome sequence is to study its end products, the proteins (Miernyk et al., 2011). Moreover, RNA and protein profiling technologies have recently been applied in parallel to improve protein identification in proteomic studies (Desgagne-Penix et al., 2010; Lundberg et al., 2010). This has led to an emerging field of biological research at the intersection of proteomics and genomics referred to as proteogenomics, which can be used to either refine genome annotation in order to identify novel translated products or to assign and identify more spectra and, therefore, identify more proteins (Ansong et al., 2008). During the last years novel sequencing technologies, such as RNA-seq, besides high-throughput MS-based proteomics have sped-up proteogenomic research (Helmy et al., 2012). However, there is no available public web resource for mining the genomic and transcriptomic data of fern (Aya et al., 2015).

The goal of the present study is to create an extensive protein resource for the gametophyte of D. affinis ssp. affinis that will be used to gain insights into the molecular basis of apogamy. Our proteogenomic approach, using a species lacking an annotated genome, increased four times the number of indentified peptides as compared to using only publicly available data bases and allowed us to identify a total of 1,397 protein clusters with 5,865 unique peptide sequences. All the raw RNA sequencing files in fastq format and the de novo transcriptome assembly in fasta format have been deposited at the European Nucleotide Archive (ENA), accession number PRJEB18522, and all proteomics raw data and the relevant derived files have been deposited at ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005423.

MATERIALS AND METHODS

In vitro Culture of Spores and Gametophytes

Spores of D. affinis ssp. affinis obtained from sporophytes growing in the forest of Turón (Asturias, Spain) were soaked in water for 2 h and then washed for 10 min with a solution of NaClO (0.5%) containing Tween 20 (0.1%). Then, they were rinsed three times with sterile distilled water. Spores were centrifuged at 1,300 g for 3 min between rinses, and then cultured in 500-ml Erlenmeyer flasks containing 100 mL of Murashige and Skoog (MS) medium (Murashige and Skoog, 1962), supplemented with 2% sucrose (w/v), pH 5.7.
Gametophytes at three developmental stages—filamentous, spatula, and heart (in the last stage with visible signs of an evolving apogamic center)—were collected to carry out the molecular analyses (Figure 2). Cultures of filamentous gametophytes were obtained by maintaining the spores in liquid cultures placed on a gyratory shaker (75 rpm) for 50 days. Cultures of spatula and heart stage gametophytes were cultured in Petri dishes with 25 mL of MS medium containing 2% sucrose (w/v) and 0.7% agar, pH 5.7, for 65 days. All cultures were maintained at 25°C under cool-white fluorescent light (40 \( \mu \text{mol m}^{-2} \text{s}^{-1} \)) with a 16-h photoperiod. For RNA extraction, 100 mg of fresh plant material was weighed, immediately frozen in liquid nitrogen, and kept at −80°C until use. For proteomic analyses, gametophytes were lyophilized and kept at −20°C until use. Three biological replicates were used for RNA sequencing and two biological replicates were used for proteomics.

**RNA Extraction**

Plant material, 100 mg of gametophytes at specific stages, was homogenized by adding glass beads to an Eppendorf tube and shaking with a Silamat S5 shaker (Ivoclar Vivadent, Schaan, Liechtenstein) twice during 10 and 5 s, respectively. Total RNA was isolated using the Spectrum\textsuperscript{TM} Plant Total RNA kit (Sigma-Aldrich, Buchs, Switzerland). DNA was removed using the TURBO DNA-free kit (Life Technologies, Carlsbad, CA), and checked to determine quality using the Bioanalyzer Agilent RNA 6000 Pico Kit (Agilent Technologies, Waldbronn, Germany).

**Library Preparation**

The quality of the isolated RNA was determined with a Qubit\textsuperscript{®} (1.0) Fluorometer (Life Technologies, California, USA) and a Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany). Only those samples with a 260/280 nm ratio between 1.8–2.1 and a 28/18S ratio within 1.5–2.0 were further processed. The TruSeq RNA Sample Prep Kit v2 (Illumina, San Diego, CA) was used in successive steps. Briefly, total RNA samples (100–1,000 ng) were enriched for polyA RNA and then reverse-transcribed into double-stranded cDNA. The cDNA samples were fragmented, end-repaired, and polyadenylated before ligation of TruSeq adapters (Table S1) containing the index for multiplexing. The products resulted in a smear with an average fragment size of approximately 260 bp. The libraries were normalized to 10 nM in Tris-Cl 10 mM, pH8.5, with 0.1% Tween 20.
Cluster Generation, Sequencing, De Novo Assembly, Transcriptome Coverage, and Data Quality

Each of the six samples (filamentous and heart tissues, three samples each) was sequenced on the Illumina HiSeq 2000 employing a 2 × 100 bp protocol. The number of raw reads generated was in the range 70–92 M. The fastq files were preprocessed using far (https://wiki.gacrc.uga.edu/wiki/FAR), the predecessor to flexbar (http://sourceforge.net/projects/flexbar/). The minimum length was set to 50 bp and adapters were trimmed as long as they would overlap 5 bases with the read. The reads passing these filters were then joined using fastqjoin (https://pods.iplantcollaborative.org/wiki/display/DEapps/Fastq-Join) so to maximize the length of the reads prior to the transcriptome assembly.

The joined reads were then passed onto Trinity (version 2013-02-25, http://trinityrnaseq.sourceforge.net) for the de novo transcriptome assembly with default settings. The total number of putative transcripts generated by Trinity was 436,707.

Relative abundances of the transcripts originating from the different samples were estimated using RSEM (http://www.biomedcentral.com/1471-2105/12/323) by mapping to the newly generated transcriptome and differential expression, both at isoform and gene level, was measured with EBseq (http://www.biostat.wisc.edu/~kendzior/EBSEQ/).

The putative 436,707 transcripts as generated by Trinity were 6-frame translated using six pack (http://emboss.sourceforge.net/apps/release/6.6/emboss/apps/sixpack.html), 330,049 amino acid (AA) sequences longer than 60 AA were kept in the NGS database (DB). To add some minimal annotation to our NGS DB sequences, each was blasted (blastp) against the Swissprot DB, a well curated multi-species database where most of the proteins have an associated function. The description line of the corresponding SSTDB entry was extended if the best scoring BLAST hit was found with an e-value of 1E-4 or smaller. This cross species annotation of the closest BLAST hit should be seen dynamic (while the actual sequences are rather static): since databases get better curated overtime, there might be better homologs to annotate our sequences in the future.

Protein Extraction

From each of the four samples (filamentous and heart tissues, two samples each) an amount of 20 mg dry weight of plant gametophytes were homogenized using a Silamat S5 shaker (Ivoclar Vivadent, Schaan, Liechtenstein). Homogenized samples were solubilized in 800 μL of buffer A [0.5 M Tris-HCL pH 8.0, 5 mM EDTA, 0.1 M Hepes-KOH, 4 mM DTT,
15 mM EGTA, 1 mM PMSF, 0.5% PVP and 1 × protease inhibitor cocktail [Roche, Rotkreuz, Switzerland]) using a Potter homogenizer (Thermo Fisher Scientific, Bremen, Germany). Proteins were extracted in two steps: first, the homogenate was subjected to centrifugation at 16,200 g for 10 min at 4°C on a tabletop centrifuge and, second, the supernatant was subjected to ultracentrifugation at 117–124 kPa (~100,000 g) for 45 min at 4°C. Post-ultracentrifugation the supernatant contained the soluble protein fraction. The pellet from the first ultracentrifugation was re-dissolved in 200 µL of buffer B (40 mM Tris base, 40 mM DTT, 4% SDS, 1 × protease inhibitor cocktail [Roche, Rotkreuz, Switzerland]) to extract membrane proteins using the ultracentrifuge as described before. The supernatant after the second ultracentrifugation step contained the membrane protein fraction. Ultracentrifugation was performed using an Airfuge (Beckman Coulter, Pasadena, CA). Protein concentrations were determined using a Qubit fluorometer (Invitrogen, Carlsbad, CA).

1D Gel Electrophoresis

Approximately 1 mg protein per each soluble and membrane fraction was loaded separately onto a 0.75 mm thick, 12% SDS-PAGE mini-gel. Samples were treated with sample loading buffer and 2 M DTT, heated at 99°C for 5 min, followed by a short cooling period on ice, and loaded onto the gel. 1D gel electrophoresis was performed at 150 V and 250 mA for 1 h in 1X Running Buffer.

Protein Separation and In-Gel Digestion

After 1D SDS-PAGE each gel lane was cut into six 0.4 cm wide sections using a custom-made gel cutter, resulting in 48 slices. These slices were further fragmented into smaller pieces and subjected to 10 mM DTT (in 25 mM AmBic pH8) for 45 min at 56°C and 50 mM Iodoacetamide for 1 h at RT in the dark prior to trypsin digestion at 37°C overnight (Baerenfaller et al., 2008). The small pieces were washed twice with 100 µL of 100 mM NH₄HCO₃/50% acetonitrile, and washed once with 50 µL acetonitrile. All three supernatants were discarded and peptides digested with 20 µL trypsin (5 ng/µL in 10 mM Tris/2 mM CaCl₂, pH 8.2) and 50 µL buffer (10 mM Tris/2 mM CaCl₂, pH 8.2). After microwave-heating for 30 min at 60°C, the supernatant was removed and gel pieces extracted once with 150 µL 0.1% TFA/50% acetonitrile. All supernatants were combined and dried, and samples were then dissolved in 15 µL 0.1% formic acid/3% acetonitrile and transferred to auto-sampler vials for liquid chromatography (LC)-MS/MS where 5 µL were injected.

Mass Spectrometry and Peptide Identification (Orbitrap XL)

The samples were analyzed on a LTQ Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) coupled to an Eksigent Nano HPLC system (Eksigent Technologies, Dublin, CA). Solvent composition of buffer A was 0.2% formic acid/1% acetonitrile, and of buffer B 0.2% formic acid/99.8% acetonitrile. Samples were dissolved in 3% acetonitrile/0.1% formic acid. Peptides were loaded onto a self-made tip column (75 µm × 80 mm) packed with reverse phase C18 material (AQ, particle size 3 µm, 200 Å) (Bischoff GmbH, Leonberg, Germany) and eluted at a flow rate of 200 nL per min. The following LC gradient was applied: 0 min: 5% buffer B, 56 min: 40% B, 60 min: 47% B, 64 min: 97% B, 71 min: 97% B. Mass spectra were acquired in the m/z range 300–2000 in the Orbitrap mass analyzer at a resolution of 60,000 at m/z 400. MS/MS spectra were acquired in a data-dependent manner from the five most intense signals in the ion trap, using 28% normalized collision energy and an activation time of 30 ms. The precursor mass isolation width was set to m/z 3.0. Charge state screening was enabled, and singly charged precursor ions and ions with undefined charge states were excluded. Precursor masses already selected for MS/MS acquisition were excluded from further selection for 120 s. MS/MS spectra were converted to the Mascot generic format (.mgf) using MascotDistiller 2.3.2 and the parameters recommended for Orbitrap instruments. These .mgf files were submitted to Mascot (Matrix Science, London UK; version 2.4.01) for searching. Trypsin was selected as the proteolytic enzyme. Mascot was set up to search against the in-house generated SSTDB (forward entries: 330,049) combined with the publicly available VPDB (forward entries: 1,031,407, downloaded from uniprot.org in March 2012), and a set of 260 known mass spectrometry contaminants in a target-decoy strategy (using reversed protein sequences). The concatenated DB is available online (http://fgcz-r-021.uzh.ch/fasta/p1222_combo_NGS_n_Viridi_20160205.fasta). Data was searched with a fragment ion mass tolerance of ±0.6 Da and a precursor mass tolerance of ±10 ppm. A maximum of 2 missed cleavages were allowed. Carbamidomethylation of cysteine was specified as a fixed modification, and deamidation (N, Q), Gln->pyro-Glu (N-term Q), oxidation (M) were specified in Mascot as variable modifications.

Protein Identification, Verification, and Bioinformatic Downstream Analysis

Scaffold software (version Scaffold 4.2.1, Proteome Software Inc., Portland, OR) was used to validate MS/MS-based peptide and protein identifications. Mascot results were analyzed together using the MudPIT option. Peptide identifications were accepted if they scored better than 95.0% probability as specified by the PeptideProphet algorithm with delta mass correction, and protein identifications were accepted if the ProteinProphet probability was above 95%. Proteins that contained same peptides and could not be differentiated based on MS/MS alone were grouped to satisfy the principles of parsimony using scaffolds cluster analysis option. Only proteins that met the above criteria were considered positively identified for further analysis. The amount of random matches was evaluated by performing the Mascot searches against a database containing decoy entries and checking how many decoy entries (proteins or peptides) passed the applied quality filters. The peptide FDR and protein FDR was estimated at 0.21 and 0.93% respectively, indicating the stringency of the analysis. A total of 2,525 unique proteins were assembled into 1,397 protein clusters using Scaffold. The Spectrum Report from Scaffold satisfying the criteria mentioned above was exported and for each identified
peptide-spectrum-match (PSM) and each peptide, the origin of
the DB (being either from the VPDB, the SSTDB, or identified
in both DBs) was evaluated. PSMs for which more than one
hit was generated with exactly the same score but a different
peptide sequence were considered as conflicts and omitted from
subsequent analyses. These are cases where the AA composition
of the two assignments are the same but the first two or three
residues are permuted or represent Leu/Ile switches as these
are isobaric AAs. All proteomics data have been deposited to the
ProteomeXchange Consortium via the PRIDE partner repository
with the dataset identifier PXD005423 (Vizcaíno et al., 2016).

RESULTS

Using a De Novo Generated SSTDB Greatly
Improves Peptide Identification in the
Proteome of D. affinis

The gametophytic tissue of the fern D. affinis was used to generate
its proteomic profile by using LC-MS/MS. The spectra were
searched against a concatenated VPDB in addition to the new
protein DB that was created based on transcriptomic datasets
obtained in the present study (SSTDB) in order to identify PSMs
from any of the two databases. This search database is large
for a single organism and, therefore, probably redundant and
biased to an inherent problem of the proteogenomic approach
where some transcripts may not be completely assembled and,
therefore, result in shorter sequences in general. To back-up this
observation, we compared sequence lengths in SSTDB to VPDB
and other organism-specific databases (Figure S1). It can be seen
that on average the sequences in the SSTDB are clearly shorter.
The large size of the SSTDB is also a result of the six-frame
translation where usually only one of the six translations is
correct.

Because of the lack of a completely species-specific annotated
genome, we used the concatenated SSTDB and VPDB, which
increases the chance for random matching due to the large search
space generated. Thus, higher scores are required for individual
PSMs compared to searching smaller databases. Here, we did
not want to omit the full VPDB but accepted the loss of some
peptide/protein identifications.

As expected, proteins were more easily detected if they are
more abundant (assuming correlation of transcript and protein
abundance; Figure 3). The combination of both transcriptome
and proteome methodologies yielded a total 1,397 true forward
protein clusters with 5,865 unique peptide sequences identified
(protFDR 0.93%; Table S2). The strategy of searching against
an orthologue DB (VPDB) concatenated to a newly generated
protein DB derived from species-specific transcriptome data
(SSTDB) dramatically improved protein identification. Of
all uniquely identified peptide sequences, more than 77.8%
were exclusively matched in the SSTDB, while only about
15.7% were exclusively matched in the VPDB (Table 1).
The intersection of peptides identified in both DBs was
ca 6% after removing conflicting assignments. This is also
obvious at the protein cluster level: more than 1,068 clusters
(76.45%) were exclusively identified in the SSTDB, while
only 329 clusters (23.55%) would have been identified if
we had searched only against the VPDB. The intersection
revealed 167 clusters (11.95%), which leaves only 162 clusters
(11.6%) that are exclusively identified in the VPDB. This
represents about 3.8 times more peptide sequences that
could be identified using this proteogenomics approach as
compared to using VPDB alone. The overview of the full
experiment workflow is illustrated in Figure 4. A list of all
proteins identified in this study is provided in Table S2 or the
Scaffold file (.sf3), which can be downloaded from the PRIDE
repository.

![TPM expression of RNA-Seq results and the respective Protein IDs](image)

**FIGURE 3** Distribution for RNA-Seq counts (TPM) and RNA-expression values of the identified proteins in the gametophyte of Dryopteris affinis ssp. affinis. F, filamentous; H, heart stage; PrX, proteomics; TPM, transcripts per million.
Functional Annotation Reveals a High Metabolic Activity of *D. affinis* Gametophytes

To gain information about possible functions of the proteins identified within VPDB, we assigned them to *gene ontology* (GO) functional categories (“biological process,” “molecular function,” “cellular component”). Our data reveal the usual behavior in a shotgun proteomics approach, in which proteins of high abundance are predominantly identified; however, some interesting categories that emphasize the nature of the tissue under investigation were also observed.

Under “biological process” the GO categories include “cellular processes” and “development and differentiation” as expected for developing gametophytes. The proteome of *D. affinis* gametophytes is dominated by processes that indicate a high metabolic activity. In addition, proteins involved in “regulation,” “defense,” “response to stimulus,” and “signalizing,” reflect the intensive interactions of free living gametophytes with their environment.

Under “molecular functions” three GO categories dominate, namely “ion binding,” “enzyme activity,” and “nucleotide binding,” while under “cellular components” we mostly found proteins localized to “plastids” and “cytoplasm,” but also to the “nucleus” and “membrane” compartments. Proteins from virtually all cellular compartments as well as the extracellular cell wall were identified (data not shown).

Finally, we also identified proteins without a GO annotation, among others the Coiled-coil domain-containing protein 18, Elicitor-responsive protein 3, GEM-like protein 1, LEA1, UPF0763 protein NAMH 0545, and the B2 protein.

**D. affinis** Gametophytes Contain Proteins with Similarity to Plant, Animal, and Fungal Proteins

More than half of all identified *D. affinis* proteins had BLAST hits to proteins from higher plants, followed by hits from animals, not mapped entries, and lower plants and algae as the most abundant (Figure 5). Table 2 shows the best species match for proteins identified within the VPDB or the species used to extend the description with useful annotation from SSTDB for up to a cumulative 70% of all identified proteins. For proteins identified within the SSTDB, we indicate the category and e-value for the BLAST annotation. Interestingly, for the identified proteins most of the BLAST hits were found with small e-values (e < 1E-20). In contrast to the complete database where most of the BLAST hits were found with an e-value above 1E-6 (Table S3).

Most hits had similarity to proteins encoded by the best-annotated genomes of higher plants, namely *Arabidopsis thaliana* (mouse ear cress) and the monocot *Oryza sativa* (rice; Table 2). However, there might be a bias here because, as the best annotated plant species, those are the ones with most entries in the swissprot DB. Surprisingly, in an identification based on the SSTDB entry instead of the VPDB entry, they were not followed by other plants with well-annotated genomes, including *Solanum lycopersicum* (tomato) and *Vitis vinifera* (grape), but rather by *Homo sapiens* (human), *Bos taurus* (cattle), and *Mus musculus* (mouse). Apart from these animals and several additional plant species, hits were also identified to proteins from the protozoan *Dictyostelium discoideum* (slime mold) and the fungus *Schizosaccharomyces pombe* (fission yeast; Table 2).

Figure S2 shows pairwise alignments for proteins discussed here, which were identified within the SSTDB and had an annotation from BLAST. Figure S3 provides annotated PSMs for proteins for which the basis of identification is a single confident peptide sequence.

Using the Scaffold software and the file provided in the PRIDE repository, GO categories can be visualized for each protein or also compared across samples, and blastp searches can directly be launched at the NCBI homepage.

**DISCUSSION**

Plant reproduction is key to understanding plant development but our knowledge on the molecular basis behind asexual reproduction or apomictic developmental programs is scarce. Ferns are frequent apogamous species and as such they can provide valuable information. Studies with an “omics” approach are scarce in ferns due to their complex, large genomes and low agronomic value (Bona et al., 2010; Der et al., 2011; Cordle et al., 2012; Shen et al., 2014; Aya et al., 2015; de Vries et al., 2015). This paper reports the first protein resource for a fern gametophyte, namely the apogamous gametophyte of *D. affinis* ssp. *affinis*. Although no genome sequence is yet available for this non-model species, it could prove useful for future research into the basic principles of apogamy, a process of great importance to agriculture (Spillane et al., 2004).

**Proteogenomics is a Powerful Approach to Identify Proteins in Proteomic Studies of Non-Model Species**

Identifying peptides and proteins from non-sequenced organisms has already been examined before. This is always possible based on completely identical peptide sequences...
between the species under investigation and the species in the search database. This only becomes problematic if the species under investigation is very distantly related to species where protein sequences are available in the search database. In these cases, peptide and protein identification can be performed by estimating the quality of a tandem mass spectrum, and if the quality is sufficient, de-novo sequencing followed by MS homology searching (Siddique et al., 2006; Grossmann et al., 2007; Vertommen et al., 2011). The major advantage of first generating a SSTDB is usually the increased sensitivity in the number of protein identifications as well as the number of peptides identified per protein.

In this study, we identified four-times more peptides with high confidence using a SSTDB concatenated with the VPDB than with the public VPDB alone. Although the concatenation of these databases results in a very large database with many homologous entries, our results demonstrate that the combination of proteomic and transcriptomic resources is essential to make adequate biological interpretations. In agreement with previous studies, we show that the sole use of the VPDB—or any other publicly available database for protein identification—is inefficient in non-model species, since they are under-represented in most databases, resulting in poor identification rates (Romero-Rodríguez et al., 2014).

As a result of searching SSTDB concatenated with the VPDB, we could identify about 1,400 protein clusters from gametophytic tissue of D. affinis. According to their assigned GO category under “biological process,” and according to the functions of mapped orthologous proteins, many proteins are associated with a high metabolic activity in agreement with the free-living nature of...
fern gametophytes that are photosynthetically active and thus autotroph (Der et al., 2011; Cordle et al., 2012). Similar to what was found in the transcriptomes of the MMC and female gametophyte of the flowering plant *A. thaliana* (Wuest et al., 2010), proteins involved in RNA metabolism and translation also feature prominently in the *D. affinis* proteome.

**D. affinis** Proteins are Homologs to Proteins Involved in Reproduction of Higher Plants

Among the identified proteins, those related to the biology of fern gametophytes are of special relevance to understand apogamy and the molecular basis of asexual reproduction (*Table 3*). As a reproductive structure, the gametophyte of ferns could be expected to be equivalent to the tissues giving rise to male (pollen) and female gametophytes (embryo sacs) in flowering plants. In line with this, *in silico* expression of the apogamy library *Arabidopsis* homologs, enriched in flower and seed structures, was reported for the apogamous gametophyte of *Ceratopteris richardii* (Cordle et al., 2012). Hence, despite the rapid evolution of reproductive proteins (Swanson and Vacquier, 2002), we found several homologs of proteins implicated in the reproduction of higher plants in the proteome of *D. affinis* gametophytes. In fact, many of the genes involved in development of the flower, for example, have homologs in non-flowering clades, illustrating the importance of examining the basic biology of taxa other than model organisms (Hasebe, 1999). Several proteins identified from the apogamous gametophyte of *D. affinis* have been implicated in embryo development of higher plants (*Table 3*). Among them are members of the LATE EMBRYOGENESIS ABUNDANT (LEA) type 1 family: embryonic protein DC-8, LEA1, the zygotic DNA replication licensing factor MCM6-A, some receptor-like kinases (RLKs), and the GEM-like protein 1 (*Table 3*). RLKs, such as those of the SOMATIC EMBRYOGENESIS RECEPTOR KINASE (SERK) subfamily, play a role in the acquisition of embryogenic competence (Hecht et al., 2001; Albertini et al., 2005). We also identified homologs of the leucine-rich repeat (LRR)-RLK GASSHO1, which exhibits uniform expression in the embryo from the globular to the mature stage (Tsuwamoto et al., 2008; *Table 3*). In addition, proteins involved in plant reproduction were identified, such as the pollen-expressed RLK ligand LAT52 (Tang et al., 2002) and the ubiquitin receptor DA1, controlling seed and organ size through the maternal sporophyte by restricting the period of cell proliferation (Li et al., 2008; *Table 3*). Furthermore, we identified homologs of animal proteins, such as Janus-B, which regulates somatic sex differentiation in *Drosophila melanogaster* (fruit fly; Yanicostas et al., 1989) and Radial Spoke Head 1 (RSPH1), required for sperm motility in humans (*Table 3*; Onoufriadis et al., 2014).

**Identification of Proteins with a Potential Role in Apogamy**

In both apomixis and apogamy, unreduced cells form an embryo without fertilization and, thus, they share some common features. Moreover, the mechanism of asexual reproduction in lower and higher plants appears to be controlled by overlapping sets of genes (Cordle et al., 2012). ARGONAUTE (AGO) proteins play important roles in RNA-mediated silencing during plant development, including reproduction (Olmedo-Monfil et al., 2014). In this study, we identified a fern protein homologous to ARGONAUTE10/PINHEAD/ZWILLE (AGO10; *Table 3*), which represses cell entry into sexual reproduction and contributes...
### TABLE 2 | Best matching species for all the proteins identified and the respective e-value category if the identification was identified with the SSTDB and having a blastp homolog.

| Species/E-value category | E-value < 1E-20 (%) | 1E-20 < E-value < 1E-10 | 1E-10 < E-value < 1E-6 (%) | 1E-6 < E-value (%) | No_useful_blastp_hit | Viridi_entry (%) | Grand total (%) | Cumulative total (%) |
|--------------------------|---------------------|--------------------------|----------------------------|-------------------|----------------------|-----------------|------------------|---------------------|
| Arabidopsis thaliana     | 94.75               | 577                      | 0.66                       | 4                 | 0.49                 | 3               | 0.49             | 3                   |
| NA: Not mapped with blastp |                     |                           |                            |                   |                      |                 |                  |                     |
| Oryza sativa             | 71.76               | 61                       | 1.64                       | 1                 |                      |                 |                  |                     |
| Physcomitrella patens    | 81.08               | 30                       | 5.41                       | 2                 | 8.11                 | 3               | 5.41             | 2                   |
| Homo sapiens             | 2.7                 | 1                        |                            |                   |                      |                 |                  |                     |
| Populus trichocarpa      |                      |                           |                            |                   |                      |                 |                  |                     |
| Selaginella moellendorff |                      |                           |                            |                   |                      |                 |                  |                     |
| Vitis vinifera           | 5.14                | 2                        |                            |                   |                      |                 |                  |                     |
| Bos taurus               | 96.53               | 26                       | 3.33                       | 1                 | 3.33                 | 1               |                  |                     |
| Mus musculus             | 60                  | 18                       | 16.67                      | 5                 | 13.33                | 4               | 10               | 3                   |
| Glycine max              | 64.29               | 18                       |                            |                   |                      |                 |                  |                     |
| Pisum sativum            | 62.96               | 17                       | 22.22                      | 6                 | 7.41                 | 2               |                  |                     |
| Solanum lycopersicum     | 69.23               | 18                       | 7.69                       | 2                 |                      |                 |                  |                     |
| Ploa stichensis          |                      |                           |                            |                   |                      |                 |                  |                     |
| Zea mays                 | 65.22               | 15                       | 4.55                       | 1                 | 9.09                 | 2               | 4.55             | 1                   |
| Dicyostelium discoideum  | 81.82               | 18                       | 1.67                       | 5                 | 13.33                | 4               | 10               | 3                   |
| Nicotiana tabacum        | 80                  | 16                       |                            |                   |                      |                 |                  |                     |
| Medicago truncatula      |                      |                           |                            |                   |                      |                 |                  |                     |
| Solanum tuberosum        | 89.47               | 17                       |                            |                   |                      |                 |                  |                     |
| Hordeum vulgare          |                      |                           |                            |                   |                      |                 |                  |                     |
| Spinacia oleracea        | 70.59               | 12                       |                            |                   |                      |                 |                  |                     |
| Ricinus communis         | 6.25                | 1                        |                            |                   |                      |                 |                  |                     |
| Schizosaccharomyces      | 66.67               | 10                       | 13.33                      | 2                 | 13.33                | 2               | 6.67             | 1                   |
| Sorghum bicolor          | 6.67                | 1                        |                            |                   |                      |                 |                  |                     |
| Adiantum capillus-veneris| 75                  | 9                        |                            |                   |                      |                 |                  |                     |
| Acanthochloa marina      | 100                 | 11                       |                            |                   |                      |                 |                  |                     |
| Arabidopsis lyrata       |                      |                           |                            |                   |                      |                 |                  |                     |
| Bacillus subtilis        | 60                  | 6                        |                            |                   |                      |                 |                  |                     |
| Rattus norvegicus        | 60                  | 6                        |                            |                   |                      |                 |                  |                     |
| Daucus carota            | 44.44               | 4                        |                            |                   |                      |                 |                  |                     |
| Triticum aestivum        | 55.56               | 5                        |                            |                   |                      |                 |                  |                     |
| Brassica napus           | 75                  | 6                        |                            |                   |                      |                 |                  |                     |
| Caenorhabditis elegans   | 87.5                | 7                        |                            |                   |                      |                 |                  |                     |
| Chlamydomonas reinhardtii| 12.5                | 1                        |                            |                   |                      |                 |                  |                     |
| Chlorella variabilis     |                      |                           |                            |                   |                      |                 |                  |                     |
| Hordeum vulgare          |                      |                           |                            |                   |                      |                 |                  |                     |
| Micromonas sp            |                      |                           |                            |                   |                      |                 |                  |                     |

(Continued)
| Species/E-value category | E-value < 1E-20 (%) | 1E-20 < E-value < 1E-10 | 1E-10 < E-value < 1E-6 (%) | 1E-6 < E-value (%) | No_useful_blastp_hit | Viridi_entry (%) | Grand total (%) | Cumulative total (%) |
|--------------------------|---------------------|--------------------------|-----------------------------|-------------------|---------------------|-----------------|-----------------|---------------------|
| Xenopus laevis           | 75                  | 6                        | 25                          | 2                 | 0.35                | 8               | 63.96           |                     |
| Cicer arietinum          | 71.43               | 5                        |                             |                   | 31                  | 7               | 64.27           |                     |
| Danio rerio              | 57.14               | 4                        |                             |                   | 31                  | 7               | 64.58           |                     |
| Hordeum vulgare          | 71.43               | 5                        |                             |                   | 31                  | 7               | 64.89           |                     |
| Catharanthus roseus      | 83.33               | 5                        |                             |                   | 31                  | 7               | 65.15           |                     |
| Gallus gallus            | 100                 | 6                        |                             |                   | 26                  | 6               | 65.41           |                     |
| Ostreococcus lucimarinus | 83.33               | 5                        |                             |                   | 26                  | 6               | 65.67           |                     |
| Petunia hybrida          | 100                 | 6                        |                             |                   | 26                  | 6               | 65.93           |                     |
| Picea abies              | 100                 | 6                        |                             |                   | 26                  | 6               | 66.19           |                     |
| Volvox carteri           | 66.67               | 4                        |                             |                   | 26                  | 6               | 66.45           |                     |
| Cucumis sativus          | 80                  | 4                        | 20                          | 1                 | 22                  | 5               | 66.67           |                     |
| Drosophila melanogaster  | 60                  | 3                        | 20                          | 1                 | 22                  | 5               | 66.89           |                     |
| Escherichia coli         | 80                  | 4                        | 20                          | 1                 | 22                  | 5               | 67.11           |                     |
| Gossypium hirtum         | 20                  | 1                        |                             |                   | 22                  | 5               | 67.33           |                     |
| Nicotiana plumbaginifolia| 80                  | 4                        |                             |                   | 22                  | 5               | 67.55           |                     |
| Saccharomyces cerevisiae | 60                  | 3                        | 40                          | 2                 | 22                  | 5               | 67.77           |                     |
| Anemia phyllitis         | 75                  | 3                        |                             |                   | 22                  | 5               | 67.94           |                     |
| Angiopteris evecta       | 25                  | 1                        |                             |                   | 22                  | 5               | 68.11           |                     |
| Asparagus officinalis    | 50                  | 2                        |                             |                   | 22                  | 5               | 68.28           |                     |
| Beta vulgaris            | 75                  | 3                        |                             |                   | 22                  | 5               | 68.45           |                     |
| Nostoc sp.               | 100                 | 4                        |                             |                   | 22                  | 5               | 68.62           |                     |
| Polystichum munitum      | 75                  | 3                        |                             |                   | 22                  | 5               | 68.79           |                     |
| Pseudomonas putida       | 50                  | 2                        |                             |                   | 22                  | 5               | 68.96           |                     |
| Pteris vittata           | 100                 | 4                        |                             |                   | 22                  | 5               | 69.13           |                     |
| Ricinus communis         | 100                 | 4                        |                             |                   | 17                  | 4               | 69.3            |                     |
| Synechocystis sp.        | 75                  | 3                        | 25                          | 1                 | 17                  | 4               | 69.47           |                     |
| Thermosynechococcus      | 100                 | 4                        |                             |                   | 17                  | 4               | 69.64           |                     |
| Xenopus tropicalis       | 75                  | 3                        | 25                          | 1                 | 17                  | 4               | 69.81           |                     |
| Aedes aegypti            | 100                 | 3                        |                             |                   | 17                  | 4               | 69.94           |                     |
TABLE 3 | List of selected proteins from those identified in the gametophyte of *Dryopteris affinis* ssp. *affinis* by transcriptome (SSTDB) or proteome (VPDB) databases.

| Category                        | # Proteins in cluster | Index in supp table | Protein accession | Origin of protein | Best matching swissprot accession | Gene name | Description | eValue (blastp) | Unique identified peptide sequences | %Coverage |
|---------------------------------|-----------------------|---------------------|-------------------|-------------------|-----------------------------------|-----------|-------------|--------------|-----------------------------------|-----------|
| Cell wall modifications         | 2                     | 578                 | 313076-61_2_ORF2  | SSTDB             | Q92T66                            | E134_MAIZE | Endo-1,3;1,4-beta-D-glucanase | 1.00E-42 | 3                               | 15.10     |
| Cell wall modifications         | 3                     | 693                 | 18361-579_1_ORF2(+) | SSTDB             | Q8VY23                            | PME53_ARATH | Probable Pectinesterase 53    | 3.00E-119 | 2                               | 9.02      |
| Embryo development              | 1                     | 134                 | 306456-65_2_ORF1[2] | SSTDB             | P20075                            | LEAD8_DAURA | Embryonic protein DC-8       | 2.00E-20 | 11                              | 18.50     |
| Embryo development              | 1                     | 794                 | 378787-31_3_ORF2  | SSTDB             | Q9SE96                            | GEM1_ARATH  | GEM-like protein 1           | 6.00E-82 | 2                               | 7.59      |
| Embryo development              | 4                     | 188                 | 363021-38_2_ORF1  | SSTDB             | O0LQO5                            | GSO1_ARATH  | LRR-receptor-like serine/threonine-protein kinase GSO1 | 2.00E-99 | 10                              | 15.70     |
| Embryo development              | 1                     | 722                 | 245753-102_2_ORF1 | SSTDB             | Q498J7                            | MOZI_ARATH  | Zygotic DNA replication licensing factor mcm6-A | 0         | 3                               | 4.44      |
| Phytohormone signaling          | 1                     | 840                 | 390256-26_3_ORF2  | SSTDB             | Q9LJ00                            | AB19B_ARATH | ABC transporter B family member 19 | 0         | 1                               | 2.63      |
| Phytohormone signaling          | 2                     | 692                 | 24435-507_4_ORF2  | SSTDB             | P48417                            | CP74_ARATH  | Allene oxide synthase, chloroplastic | 1.00E-146 | 5                               | 10.70     |
| Phytohormone signaling          | 1                     | 866                 | 254191-97_6_ORF2(+) | SSTDB             | Q99U6                             | 1A11_ARATH  | 1-aminocyclopropane-1-carboxylate synthase 11 | 1.00E-87 | 2                               | 6.58      |
| Phytohormone signaling          | 1                     | 1158                | 106730-241_3_ORF2 | SSTDB             | F4JSZ5                            | BIG1_ARATH  | Brefeldin A-inhibited guanine nucleotide-exchange protein 1 | 0         | 1                               | 0.73      |
| Phytohormone signaling          | 1                     | 762                 | 13447-660_2_ORF1(+) | SSTDB             | Q9SCZ5                            | AOCH_ARATH  | Allene oxide cyclo 4, chloroplastic | 3.00E-70 | 1                               | 4.38      |
| Phytohormone signaling          | 6                     | 356                 | 298264-69_1_ORF1[8] | SSTDB             | Q9SK82                            | UBSA1_ARATH | UDP-glycyltransferase 85A1 | 3.00E-73 | 5                               | 10.00     |
| Phytohormone signaling          | 1                     | 902                 | 30962-455_2_ORF1 (+39) | SSTDB             | Q92TR1                            | SPD1_FEA     | Spermidine synthase 1 | 0         | 1                               | 4.36      |
| Phytohormone signaling          | 2                     | 1109                | 158601-182_4_ORF1(+) | SSTDB             | Q8TL44                            | TRB2_ARATH  | Tryptophan synthase beta chain 1 | 4.00E-170 | 1                               | 2.37      |
| Phytohormone signaling          | 1                     | 671                 | 311894-62_3_ORF2  | SSTDB             | Q8CU7                             | ZEP_ARATH   | Zeaxanthin epoxidase, chloroplastic | 3.00E-32 | 3                               | 9.26      |
| Reproduction                    | 1                     | 1272                | 41629-395_1_ORF1(+) | SSTDB             | P0C7Q8                            | DA1_ARATH   | Protein DA1 | 0         | 1                               | 4.00      |
| Reproduction                    | 1                     | 1083                | 43959-384_1_ORF2  | SSTDB             | P54365                            | JANB_ARATH  | Sex-regulated protein janus-B | 7.00E-07 | 1                               | 6.91      |
| Reproduction                    | 1                     | 737                 | 133239-212_4_ORF1(+) | SSTDB             | P13447                            | LAT52_ARATH | Anthser-specific protein LAT52 | 9.00E-14 | 2                               | 7.25      |
| Reproduction                    | 1                     | 537                 | 33857-49_3_ORF1   | SSTDB             | Q6VTH5                            | RSPH_ARATH  | Radial spoke head 1 homolog | 2.00E-09 | 2                               | 3.40      |
| Reproduction(Apogamy)           | 1                     | 1202                | 429495-5_4_ORF2   | SSTDB             | Q9UGW1                            | AGO10_ARATH | Protein argonaute 10 | 3.00E-161 | 1                               | 1.85      |
| Reproduction(Apomixis)          | 1                     | 967                 | 284827-77_1_ORF2  | SSTDB             | Q9ZVD0                            | SRPT_ARATH  | Serrate RNA effector molecule | 0         | 1                               | 2.89      |
| Reproduction(Apomixis)          | 8                     | 373                 | 319623-58_1_ORF1  | SSTDB             | Q42798                            | CO31_ARATH  | Cytochrome P450 90A1 | 5.00E-129 | 3                               | 7.56      |
| Reproduction(Apomixis)          | 7                     | 751                 | 101113-248_5_ORF2(+) | SSTDB             | P42825                            | DNAJ_ARATH  | Chaperone protein dnaJ 2 | 0.00E+00 | 1                               | 2.08      |
| Reproduction(Apomixis)          | 2                     | 1230                | 65050-256_3_ORF2  | SSTDB             | Q1PGK0                            | FBL3_ARATH  | F-box/LRR-reapeat protein At1g55660 | 8.00E-06 | 2                               | 6.67      |
| Reproduction(Apomixis)          | 1                     | 1359                | 200671-137_4_ORF2(+) | SSTDB             | P92963                            | RAB1C_ARATH | Ras-related protein RABB1c | 3.00E-133 | 1                               | 6.98      |
| Reproduction(Apomixis)          | 1                     | 1301                | 162652-176_2_ORF2 | SSTDB             | Q9UV6                             | RLK90_ARATH | Probable inactive receptor kinase | 0         | 1                               | 2.22      |

(Continued)
TABLE 3 | Continued

| Category                        | # Proteins in cluster | Index in supp table | Protein accession | Origin of protein | Best matching swissprot accession | Gene name          | Description                                      | eValue (blastp) | Unique identified peptide sequences | %Coverage |
|---------------------------------|------------------------|---------------------|-------------------|------------------|----------------------------------|-------------------|-------------------------------------------------|----------------|-------------------------------------|-----------|
| Stress response                 | 2                      | 490                 | 34681-431_3_ORF2  | SSTDB            | Q06850                           | CDPK1_ARATH       | Calcium-dependent protein kinase 1               | 0.00E+00       | 1                                   | 3.16      |
| Stress response                 | 1                      | 241                 | 82340-277_1_ORF2  | SSTDB            | Q9AR14                           | Pip15_MA2E        | Aquaporin Pip1-5                                | 1.00E-173      | 4                                   | 21.50     |
| Stress response                 | 1                      | 1373                | 196957-581_1_ORF2 (1) | SSTDB          | P4659                           | LEA14_SOYBN       | Desiccation protectant protein LEA14 homolog     | 9.00E-24       | 1                                   | 3.37      |
| Stress response                 | 1                      | 1010                | 163213-176_3_ORF1 (1) | SSTDB          | P0CW97                           | POP3_ARATH        | Protein PLANT CADMIUM RESISTANCE 3              | 1.00E-36       | 3                                   | 17.70     |
| Stress response                 | 1                      | 837                 | tr|Q5IDA4|Q5IDA4_PINTA | VPDB                | NA                               | Cluster of Cinnamate 4-hydroxylase (Fragment)     | NA             | 1                                   | 5.76      |
| Stress response                 | 2                      | 449                 | 272418-85_1_ORF2  | SSTDB            | P22242                           | DRPE_ARAPL        | Desiccation-related protein                    | 7.00E-96       | 2                                   | 13.30     |
| Stress response                 | 10                     | 1087                | 213971-126_6_ORF1 | SSTDB            | P13240                           | DR206_PEA         | Disease resistance response protein 206         | 3.00E-23       | 2                                   | 9.60      |
| Stress response                 | 2                      | 703                 | 3863-1099_6_ORF1 (1) | SSTDB          | P42761                           | GSTFA_ARATH       | Glutathione S-transferase F10                   | 3.00E-70       | 1                                   | 8.89      |
| Stress response                 | 7                      | 267                 | sp|P3I082|HSP7_CUOMA | VPDB                | NA                               | Stromal 70 kDa heat shock-related protein, chloroplastic (Fragment) | NA             | 4                                   | 9.70      |
| Stress response                 | 1                      | 673                 | 2203-1350_3_ORF1  | SSTDB            | Q9SPV5                           | NEC1_NIOPL        | Nectatin-1                                       | 3.00E-30       | 3                                   | 19.70     |
| Stress response                 | 1                      | 1091                | 149244-199_3_ORF2 | SSTDB            | Q9AL48                           | CRIM1_ARATH       | Cold-regulated 413 inner membrane protein 1, Chloroplastic | 4.00E-46       | 1                                   | 4.87      |
| Stress response                 | 1                      | 941                 | 307117-64_1_ORF1 (1) | SSTDB          | A4FF33                           | VGB_SACBN         | Virginiamycin B lyase                            | 2.00E-11       | 2                                   | 3.33      |
| Stress response/Embryo development | 2                      | 320                 | 249225-100_3_ORF2 [4] | SSTDB          | Q49816                           | LEA1_CICAR        | Late embryogenesis abundant protein 1            | 5.00E-09       | 8                                   | 66.00     |

*Category* as described in the Discussion Section.
to the maintenance of the shoot apical meristem and the establishment of leaf polarity by repressing miR165/166 in A. thaliana (Liu et al., 2009). Similarly to AGO proteins, the A. thaliana SERRATE (SE) RNA effector protein, a homolog of which was also identified here (Table 3), acts as a regulator of meristem activity and leaf polarity via the miRNA pathway (Prigge and Wagner, 2001). These proteins could potentially play a role in the meristematic activity of the incipient apogamic embryo or unknown roles in the switch between sexual and asexual reproduction. We speculate that SE and some AGO family proteins may be involved in the regulation of apogamy in ferns.

Previous reports in grasses described proteins associated with apomixis and the regulation of ploidy regulation suggesting that these are inter-related phenomena (Albertini et al., 2004). In this study, we found fern homologs of some of the described proteins, such as the Ras-related proteins, 4 DNAJ domain-containing proteins, cytochrome P450, several LRR-proteins, and proteins involved in gene silencing.

Finally, an important group of proteins identified in this study play a role in cell wall modifications, including glucanasases, glycosyltransferases, and pectinases, (Table 3). Consistent with the presence of proteins associated with pectin catabolism, it has previously been reported that pectins are present in lower concentrations in ferns than in higher plants (Silva et al., 2011). Recently, gene expression in enlarging aposporous initial cells and early aposporous embryo sacs was compared to that in surrounding cells during apomictic initiation in Hiercium praetum (tall hawkweed) and, interestingly, pectinesterases and other cell wall-modifying enzymes were identified, consistent with a role of cell wall modifications in apomixis and apogamy (Li et al., 2011).

Identification of Proteins Involved in Phytohormone Signaling

The sessile lifestyle of plants requires a continuous crosstalk between the plant and its immediate environment. Phytohormones are of prime importance in this dynamic interaction to regulate and integrate overall plant growth and development. In D. affinis, auxins and gibberellins play a stimulatory role during the induction and differentiation of apogamous embryos, but phytohormones are also important for the vegetative development of the gametophyte (Menéndez et al., 2006b, 2009). In this study, we found several proteins related to the action of the classical phytohormones auxin, cytokinin, ethylene, and abscisic acid, as well as brassinosteroids, jasmonic acid, and polyamines. The above mentioned proteins may potentially participate in key aspects of vegetative and reproductive gametophyte development in ferns.

Identification of Proteins Involved in Stress Responses

A particularity of the gametophyte of ferns is its vulnerability to stress. Hence, many proteins identified in this study are related to responses to biotic or abiotic stimuli. Regarding abiotic stress, we identified several heat-shock proteins (70, 90, 105, and hsc70), the glutathione-S-transferase protein F10/EARLY RESPONSE TO DEHYDRATION13, homologs of the desiccation-related protein PCC13-62 from Craterostigma plantagineum (resurrection plant; Piatkowski et al., 1990), and many other proteins known to participate in ABA-mediated stress responses. We also found proteins involved in cellular responses to toxic substances, including PLANT CADMIUM RESISTANCE3 (Song et al., 2004). Regarding proteins involved in biotic stress responses, we identified homologs of Virginiamycin B lyase, which is involved in antibiotic resistance, Nectarin 1, which may interact with bacterial adhesins and may protect from microbial attack (Carter et al., 1999), and proteins related to the cytochrome P450 family that, in the fern species Polypodium vulgare (common polypody), are associated with ecdysteroids, which are also present in plants (phytoecdysteroids) and suggested to participate in the defense against non-adapted phytophagous invertebrates (Canals et al., 2005). It has been suggested that an increase in metabolic activity and stress responses together induce the developmental switch to apogamy (Cordlel et al., 2012). Accordingly, the in vitro conditions that induce apogamous sporophytes in angiosperms from pollen or embryo sacs, universally include a stress treatment (Shariatpanahi et al., 2006).

The combination of different “omics” approaches is a promising way to obtain a comprehensive picture of regulatory processes. By integrating reference transcriptome and proteome analyses, we greatly improved protein identification in a non-model species, providing an important basis to gain further insights into apogamy in D. affinis ssp. affinis. Studying the molecular mechanisms of asexual reproduction, i.e., the generation of clonal offspring, is an important topic aiming at the introduction of self-sustainable hybrids in agriculture. Hence, the introduction of apomixis has a tremendous potential for crop improvement, and extending our analyses to phylogenetic branches other than those of model species may help to unravel underlying processes common to a broad range of organisms.

AUTHOR CONTRIBUTIONS

HF, MC, and UG. conceived the project; HF, JG, PC, VG, and GR. performed experiments and/or analyzed data; JG, HF, and AV wrote the manuscript and contribute figures and tables; UG revised the manuscript.

ACKNOWLEDGMENTS

This research was supported by the University of Zürich, the University Research Priority Program Functional Genomics/Systems Biology, a grant from the Spanish Ministry of Education, Culture, and Sport (PR2011-0152) to promote the mobility of senior lecturers and researchers to highly qualified centers abroad, and a Prime XS project (PRIME-XS-000252), grant agreement number 262067, funded by the European Union 7th Framework Program. We thank Hans-Peter Schöb for administrative and logistic support, and Sybille Hirsch for advice on the transcriptome experiments.
SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00336/full#supplementary-material

Figure S1 | Boxplot comparison of sequence lengths for the different databases used in this study in addition to other species-specific standard databases used in proteomics.

Figure S2 | Pairwise alignments of the proteins discussed in our study with their best swissprot blastp match.

REFERENCES

Albertini, E., Marconi, G., Barcaccia, G., Raggi, L., and Falcinelli, M. (2004). Isolation of candidate genes for apomixis in Poa pratensis L. Plant Mol. Biol. 56, 879–894. doi: 10.1023/B:PLAB.0000043048-1035-9

Albertini, E., Marconi, G., Reale, L., Barcaccia, G., Porcedda, A., Ferranti, F., et al. (2005). SERK and APOSTART. Candidate genes for apomixis in Poa pratensis. Plant Physiol. 138, 2175–2195. doi: 10.1104/pp.105.062059

Ansong, C., Purvine, S. O., Adkins, J. N., Lipton, M. S., and Smith, R. D. (2008). Proteogenomics: needs and roles to be filled by proteomics in genome annotation. Brief. Funct. Genomic Proteomic 7, 50–62. doi: 10.1093/bfgp/eln010

Aya, K. K., Kobayashi, M., Tanaka, J., Ohyanagi, H., Suzuki, T., Yano, K., et al. (2015). A de novo transcriptome assembly of a fern, Lygodium japonicum, and a web resource database, Ljtrans DB.

Baerenfaller, K., Grossmann, J., Grobei, M. A., Hull, R., Hirsch-Hoffmann, Aya, K. K., Kobayashi, M., Tanaka, J., Ohyanagi, H., Suzuki, T., Yano, K., et al. (2015). de novo transcriptome assembly of a fern, Lygodium japonicum, and a web resource database, Ljtrans DB. Plant Cell Physiol. 56, e5. doi: 10.1093/pcp/pcu184

Barker, M. S., and Wolf, P. G. (2010). Unfurling fern biology in the genomics age. BioScience 60, 177–185. doi: 10.1525/bio.2010.60.3.4

Bona, E., Cattaneo, C., Cesaro, P., Marsano, F., Lingua, G., Cavallero, M., et al. (2010). Proteomic analysis of Pteris vitatta fronds: two arbuscular mycorrhizal fungi differentially modulate protein expression under arsenic contamination. Proteomics 10, 3811–3834. doi: 10.1002/pmic.200900436

Canals, D., Iruire-Santillari, J., and Casas, J. (2005). The first cytochrome P450 in ferns. Evidence for its involvement in phytoecdysteroid biosynthesis in Polypodium vulgare. FEBS J. 272, 4817–4825. doi: 10.1111/j.1742-4658.2005.04897.x

Carter, G., Graham, R. A., and Thornburg, R. W. (1999). Nectarin I is a novel, soluble germin-like protein expressed in the nectar of Nicotiana sp. Plant Mol. Biol. 41, 207–216. doi: 10.1023/A:1006353086684

Cordle, A. R., Bui, L. T., Irish, E. E., and Cheng, C. L. (2010). “Laboratory-induced apogamy and apophasy in Ceratopteris richardii,” in Working with Ferns: Issues and Applications, eds H. Fernández, A. Kumar, and M. A. Revilla (New York, NY: Dordrecht; Heidelberg; London; Springer), 25–36.

Cordle, A. R., Irish, E. E., and Cheng, C. L. (2012). Gene expression associated with apogamy commitment in Ceratopteris richardii. Sex Plant Reprod. 25, 293–304. doi: 10.1007/s00497-012-0198-z

Der, J. P., Barker, M. S., Norman, J. W., dePamphilis, C. W., and Wolf, P. G. (2011). de novo characterization of the gametophyte transcriptome in bracken fern, Pteridium aquilinum. BMC Genomics 12:99. doi: 10.1186/1471-2164-12-99

Desgagne-Penix, I., Khan, M. F., Schrier, D. C., Cram, D., Nowak, J., and Facchin, P. J. (2010). Integration of deep transcriptome and proteome analyses reveals the components of alkaloid metabolism in opium poppy cell cultures. BMC Plant Biol. 10:252. doi: 10.1186/1471-2229-10-252

de Vries, J., Fischer, A. M., Roetger, M., Rommel, S., Schluepmann, H., Brüttigam, A., et al. (2015). Cytokinin-induced promotion of root meristem size in the fern Azolla supports a shoot-like origin of euryphyllotype roots. New Phytol. 209, 705–720. doi: 10.1111/nph.13630

Fernández, H., Bertrand, A. M., and Sánchez-Tamés, R. (1996). Influence of culture conditions on apomixis in Dryopteris affinis ssp. affinis. Plant Cell Tiss. Organ Cult. 45, 93–97. doi: 10.1007/BF00043434

Fraser-Jenkins, C. R. (1986). A classification of the genus Dryopteris (Pteridophyta: Dicotyledonae). Bull. Br. Mus. 14, 183–218.

Grossmann, J., Fischer, B., Baerenfaller, K., Owiti, J., Buhmann, J. M., Gruisem, W., et al. (2007). Workflow to increase the detection rate of proteins from unsequenced organisms in high-throughput proteomics experiments. Proteomics 7, 4245–4254. doi: 10.1002/pmic.200700474

Grossniklaus, U., Nogler, G. A., and Van Dijk, P. J. (2001). How to avoid sex: the genetic control of gametophytic apomixis. Plant Cell 13, 1491–1497. doi: 10.1105/tpc.13.7.1491

Hasebe, M. (1999). Evolution of reproductive organs in land plants. J. Plant Res. 112, 463–474. doi: 10.1007/PL00013878

Hecht, V., Vielle-Calzada, J. P., Hartog, M. V., Schmidt, E. D., Boutillier, K., Grossniklaus, U., et al. (2001). The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol. 127, 803–816. doi: 10.1104/pp.010324

Helmy, M., Tomita, M., and Ishihama, Y. (2012). Peptide identification by searching large-scale tandem mass spectra against large databases: bioinformatic methods in proteogenomics. Genes Genomes Genomics 6, 76–85. doi: 10.1111/j.1365-2443.2012.01615.x

Kaźmierczak, A. (2010). “Gibberellic acid and ethylene control male sex determination and development of Anema phylitidis gametophytes,” in Working with Ferns: Issues and Applications, eds H. Fernández, A. Kumar, and M. A. Revilla (New York, NY: Dordrecht; Heidelberg; London: Springer), 49–66.

Koltunow, A. M., and Grossniklaus, U. (2003). Apomixis: a developmental perspective. Ann. Rev. Plant Biol. 54, 547–574. doi: 10.1146/annurev.arplant.54.110901.160842

Li, J. J., Liu, L., Ouyang, Y. D., and Yao, J. L. (2011). Sexual reproduction development in apomictic Eulaliopsis binata (Poaceae). Genet. Mol. Res. 10, 2326–2339. doi: 10.4238/2011.October.5.3

Li, Y., Zheng, L., Corke, F., Smith, C., and Bevan, M. W. (2008). Control of final seed and organ size by the DA1 gene family in Arabidopsis thaliana. Genes Dev. 22, 1331–1336. doi: 10.1101/gad.463608

Liu, Q., Yao, X., Pi, L., Wang, H., Cui, X., and Huang, H. (2009). The ARGONAUTE10 gene modulates shoot apical meristem maintenance and establishment of leaf polarity by repressing mirR165/166 in Arabidopsis. Plant Cell 21, 27–40. doi: 10.1105/tpc.108.063757.x

Lopez, R. A., and Renzaglia, K. S. (2014). Multiflagellated sperm cells of Ceratopteris. Am. J. Bot. 101, 2052–2061. doi: 10.3732/ajb.1400424

Lundberg, E., Fagerberg, L., Klevbrink, D., Matic, I., Geiger, T., Cox, J., et al. (2010). Defining the transcriptome and proteome in three functionally different human cell lines. Mol. Syst. Biol. 6, 450. doi: 10.1038/msb.2010.106

Manton, I. (1950). Problems of Cytology and Evolution in the Pteridophyta. New York, NY: Cambridge University Press.

Menéndez, V. F., Villacorta, N., Bernard, P., Gotor, V., Revilla, M. A., and Fernández, H. (2006a). Exogenous and endogenous growth regulators on
apogamy in Dryopteris affinis (Lowe) Fraser-Jenkins ssp. affinis. Plant Cell Rep. 25, 85–91. doi: 10.1007/s00299-005-0041-1

Menéndez, V., Revilla, M. A., Fal, M. A., and Fernández, H. (2009). The effect of cytokinins on growth and sexual organ development in the gametophyte of Blechnum spicant L. Plant Cell Tiss. Organ Cult. 96, 245–250. doi: 10.1007/s11240-008-9481-v

Menéndez, V., Revilla, M. A., and Fernández, H. (2006b). Growth and gender in the gametophyte of Blechnum spicant L. Plant Cell Tiss. Organ Cult. 86, 47–53. doi: 10.1007/s11240-006-9095-1

Miernyk, J. A., Petrova, A., Medolilla, A., Klubicova, K., Obert, B., and Hajduch, M. (2011). Using proteomic to study sexual reproduction in angiosperms. Sex Plant Reprod. 24, 9–22. doi: 10.1007/s00497-010-0149-5

Murashige, T., and Skoog, F. (1962). A revised medium for rapid growth and bioassays with tobacco culture tissues. Physiol Plant. 15, 473–497. doi: 10.1111/j.1399-3054.1962.tb08052.x

Nogler, G. A. (1984). "Gametophytic apomixis," in Embryology of Angiosperms, ed B. M. Johri (Berlin: Springer-Verlag), 475–518.

Okano, Y., Aono, N., Hiwatashi, Y., Murata, T., Nishiyama, T., Ishikawa, T., et al. (2009). A Polycomb Repressive Complex 2 gene regulates apogamy and gives evolutionary insights into early land plant evolution. Proc. Natl. Acad. Sci. U.S.A. 106, 16321–16326. doi: 10.1073/pnas.0906971106

Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., et al. (2014). Control of female gamete formation by a small RNA pathway in Arabidopsis. Nature 464, 628–632. doi: 10.1038/nature08828

Onoufiriadis, A., Shoemark, A., Schmidts, M., Patel, M., Jimenez, G., Liu, H., et al. (2014). Targeted NGS gene panel identifies mutations in RS PH1 causing pair agenesis due to radial spoke defects. J. Proteomics. 96, 245–250. doi: 10.1016/j.jprot.2013.01.027

Prigge, M. J., and Wagner, D. R. (2001). The Embryology of Angiosperms, ed H. Fernández, A. Kumar, 11–24.

Russo and Grossniklaus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Grossmann, Fernández, Chaubey, Valdés, Gagliardini, Cañal, Russo and Grossniklaus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Vizcaino, J. A., Coords, A., del-Toro, N., Díaz, H., Griss, J., Lavidas, I., et al. (2016). 2016 update of the PRIDE database and related tools. Nucleic Acids Res. 44, D447–D456. doi: 10.1093/nar/gkv1145

Whittier, D. P. (1971). The value of fers in an understanding of the alternation of generations. BioScience 21, 225–227. doi: 10.2307/1295690

Wuest, S. E., Vijverberg, K., Schmidt, A., Weiss, M., and Gheyserlinck, J., et al. (2010). Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. Curr. Biol. 20, 506–512. doi: 10.1016/j.cub.2010.01.051

Yang, H. Y., and Zhou, C. (1992). Experimental plant reproductive biology and reproductive cell manipulation in higher plants: now and the future. Am. J. Bot. 79, 354–363. doi: 10.2307/1295690

Yang, H. Y., and Zhou, C. (1992). Experimental plant reproductive biology and reproductive cell manipulation in higher plants: now and the future. Am. J. Bot. 79, 354–363. doi: 10.2307/1295690

Yanicos, C., Vincent, A., and Lepesant, J. (1989). A transcriptional and posttranscriptional regulation contributes to the sex-regulated expression of two sequence-related genes at the janus locus of Drosophila melanogaster. Mol. Cell. Biol. 9, 2526–2535. doi: 10.1128/MCB.9.6.2526