

RESEARCH ARTICLE

Deep quantitative proteomics of North American Pacific coast star tunicate (*Botryllus schlosseri*)

Dietmar Kültz¹  | Alison M. Gardell² | Anthony DeTomaso³ | Greg Stoney³ |
 Baruch Rinkevich⁴ | Yuval Rinkevich⁵ | Andy Qarri^{4,5} | Weizhen Dong¹ |
 Brenda Luu¹ | Mandy Lin¹

¹Department of Animal Sciences & Genome Center, University of California Davis, Meyer Hall, Davis, California, USA

²School of Interdisciplinary Arts and Sciences, University of Washington Tacoma, Tacoma, Washington, USA

³Department of Molecular, Cellular and Developmental Biology, University of California Santa Barbara, Goleta, California, USA

⁴Israel Oceanography & Limnological Research, National Institute of Oceanography, Haifa, Israel

⁵Helmholtz Zentrum München, Regenerative Biology and Medicine Institute, Munich, Germany

Correspondence

Dietmar Kültz, Department of Animal Sciences & Genome Center, University of California Davis, Meyer Hall, One Shields Ave., Davis, CA 95616, USA.
 Email: dkultz@ucdavis.edu

Funding information

NSF, Grant/Award Numbers: MCB-2127516, MCB-2127517; BSF, Grant/Award Number: 2021650; Division of Molecular and Cellular Biosciences, Grant/Award Numbers: 2127516, 2127517; United States-Israel Binational Science Foundation, Grant/Award Number: 2021650

Abstract

Botryllus schlosseri, is a model marine invertebrate for studying immunity, regeneration, and stress-induced evolution. Conditions for validating its predicted proteome were optimized using nanoElute® 2 deep-coverage LCMS, revealing up to 4930 protein groups and 20,984 unique peptides per sample. Spectral libraries were generated and filtered to remove interferences, low-quality transitions, and only retain proteins with >3 unique peptides. The resulting DIA assay library enabled label-free quantitation of 3426 protein groups represented by 22,593 unique peptides. Quantitative comparisons of single systems from a laboratory-raised with two field-collected populations revealed (1) a more unique proteome in the laboratory-raised population, and (2) proteins with high/low individual variabilities in each population. DNA repair/replication, ion transport, and intracellular signaling processes were distinct in laboratory-cultured colonies. Spliceosome and Wnt signaling proteins were the least variable (highly functionally constrained) in all populations. In conclusion, we present the first colonial tunicate's deep quantitative proteome analysis, identifying functional protein clusters associated with laboratory conditions, different habitats, and strong versus relaxed abundance constraints. These results empower research on *B. schlosseri* with proteomics resources and enable quantitative molecular phenotyping of changes associated with transfer from in situ to ex situ and from in vivo to in vitro culture conditions.

KEYWORDS

data-independent acquisition, ecological proteomics, evolutionary proteomics, label-free protein quantitation, marine invertebrates, nanoElute® 2, spectral library, tunicates

Abbreviations: CSI, captive spray ionization; DDA, data-dependent acquisition; DIA, data-independent acquisition.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2024 The Authors. *PROTEOMICS* published by Wiley-VCH GmbH.

1 | INTRODUCTION

B. schlosseri is a widely distributed and highly invasive, encrusting colonial sea squirt [1–4]. It represents an emerging model marine invertebrate species [1, 5–7]. Originally, native to the Mediterranean Sea and Atlantic coast of Europe, it has spread globally into shallow water temperate zones of all continents except Antarctica [4]. It is widely distributed along the east and the west coasts of the United States [4, 8–10]. The first records of its introduction to the Pacific coast of North America are from the 1940s [8]. Each colony is enclosed by a semitranslucent organic matrix (the tunic) that contains several to thousands of genetically identical units (zooids), each of which are 1–3 mm long. Zooids are arranged in structures called systems, which have a flower-like phenotype with individual zooids resembling flower petals that are connected by a ramified vascular network containing hemolymph (Figure 1) and various cell types [11, 12]. This species can produce sexually and asexually [2, 6, 13]. In sexually mature colonies, testes and ovaries are both present within the same bud, and released sperm fertilize eggs that develop into short-lived pelagic larvae [14]. Self-fertilization is minimized by the asynchronous release of sperm and eggs [15]. These larvae settle on a substrate and develop into encrusting *B. schlosseri* colonies, which can reproduce asexually by forming buds that develop into new zooids, in a weekly cycles of death and budding, known as blastogenesis. This cycle takes approximately 1 week at a temperature range of 18°C–20°C [2]. Laboratory cultures of this species are generally established by collecting gravid colonies from the field that release pelagic larvae within a few days of collection. These larvae settle on glass slides, metamorphose, and grow to adults. While the methodologies of animal maintenance differ once established [16–18], laboratory cultures are successfully propagated asexually [19–21].

Recent advancements in the fields of cancer research and cell immortalization have shed light on the fact that environmental stress accelerates the evolutionary processes of cell adaptation, favoring cellular phenotypes that outperform neighboring cells in terms of their proliferation rates [22–26]. This phenomenon serves as the foundation for spontaneous cell immortalization. However, in the realm of aquatic invertebrates, the occurrence of spontaneous cell immortalization is significantly less frequent compared to terrestrial animals or fish. In parallel, there is a rapidly growing demand for aquatic invertebrate cell lines, driven by both, basic and applied research objectives (such as for cell-based seafood production and materials for the pharmaceutical industry [27]). Therefore, revealing the changes in molecular (e.g., proteomic) networks when aquatic invertebrates are shifted from in situ to ex situ conditions or when cells from these organisms are transferred from in vivo to in vitro settings, are of considerable interest.

Here we have analyzed the proteome of laboratory-bred colonies of the star tunicate (*Botryllus schlosseri*). This species is well suited for experimental evolution studies [5, 6], at both the whole animal and cellular levels, because it can be cultured in the laboratory for many generations and multiple years, revealing unique developmental biology processes (e.g., whole body regeneration, rejuvenation torpor states, totipotent stem cells statuses) and allrecognition phenomena

Significance Statement

The star tunicate (*Botryllus schlosseri*) is an emerging model for studies on the molecular ecology of invasive species, the evolution of immunity, body regeneration, and stress-induced evolution. Despite its significance, only a handful of research facilities across the globe have managed to maintain viable laboratory-bred colonies. This species' genome was fully sequenced a decade ago, but its proteome had not been analyzed. By utilizing the new nanoElute® 2 for peptide separation, we established, for the first time, the optimal operational parameters that maximize protein identification and coverage for a colonial tunicate proteome. In addition, we characterized the first expressed proteome of any colonial tunicate in both laboratory and field contexts. Our work yielded a STRING annotated reference proteome database and a rigorously quality-filtered DIA assay library for *B. schlosseri*, which represent important new resources for comparative molecular ecology and evolutionary biology. The results include the identification of functional protein networks undergoing the most significant changes in laboratory cultures, and proteins exhibiting the highest and lowest variability across all three analyzed *B. schlosseri* populations. Since phenotypes are based on the proteome and proteins are the targets of almost all pharmaceutical drugs, proteomics provides new insights for managing this invasive species and devising strategies for the establishment of research-amenable laboratory cultures.

(fusion/rejection, natural chimerism), shedding light on the evolution of these phenomena [20, 28, 29]. This group of organisms is closely related to the vertebrates by sharing the same phylum (Chordata) [30] yet grows as flat colonial layers attached to a surface, and is one of the few marine invertebrates from which reliable primary cell cultures have been derived [12, 31].

The goal of this study was to experimentally validate predicted *B. schlosseri* proteins and to understand how the proteome is altered during the establishment of stocks of laboratory-bred colonies, which are genetically tractable and free from contamination, serving as source material for deriving primary cultures aimed at achieving immortalization [32]. This analysis involved a comparison with field-collected samples, to reveal protein networks that undergo alterations in multi-generation long-term and controlled ex situ culture conditions. The comparison included two geographically distant populations situated about 1800 km apart. Another major goal of this study was experimental validation of the first predicted *B. schlosseri* proteome and the establishment of a data-independent acquisition (DIA) assay library and STRING annotated reference database for quantitative functional proteomics studies of this species.

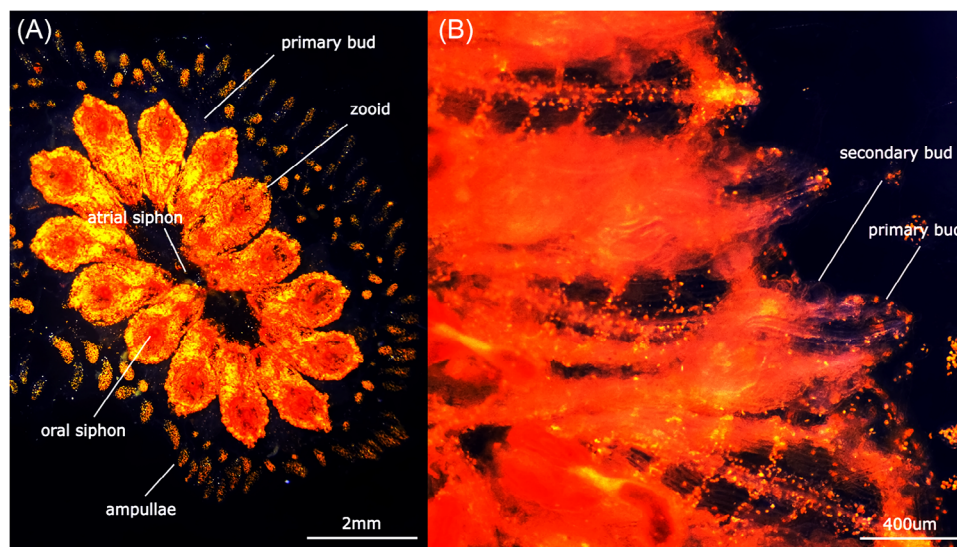


FIGURE 1 *Botryllus schlosseri* system at stage C [60, 87], enclosed by a single tunic, with dorsal (A) and ventral (B) views. At each blastogenic stage, the system comprises three generations: zooids, primary buds, and secondary buds. Zooids feature individual oral siphons, while the atrial siphon is shared across the entire system. The colony's expansion is facilitated by numerous vascular system ampullae located around the periphery.

2 | Materials and methods

2.1 | Sample preparation and experimental design

Botryllus schlosseri samples (single systems, usually consisting of up to 12 zooids, Figure 1) were collected in late summer 2022 from two marinas, one in southern California (Santa Barbara, CA, GPS coordinates: 34.414478, -119.828650) and the second in northern Washington (Des Moines, WA, GPS coordinates: 47.398657, -122.330650; 1800 km apart). Concurrently, we sampled laboratory raised *B. schlosseri* colonies that originated from gravid colonies collected from the Santa Barbara marina and were then raised for 1 year in a laboratory facility at the University of California Santa Barbara as previously described [20]. The samples were assigned to the following three groups: Santa Barbara field population (W, $n = 12$ colonies), Santa Barbara laboratory raised population (C, $n = 12$ colonies), and Des Moines field population (T, $n = 20$ colonies). All samples were snap-frozen in liquid nitrogen immediately upon harvest and stored at -80°C until processing. Samples were thawed, resuspended in lysis buffer (8 M urea buffer, 50 mM ammonium bicarbonate), and homogenized using a beadbug homogenizer and 3 mm zirconium beads (Benchmark Scientific). Extracted proteins were then reduced by 10 min incubation with 5 mM dithiothreitol at 60°C , followed by 30 min incubation with 15 mM iodoacetamide in the dark. Samples were quenched by increasing dithiothreitol to 10 mM before digestion (4 h) with trypsin/Lys-C (Thermo A41007) at a ratio of 50 parts of sample protein to 1 part of enzyme. Formic acid (5% final volume) was then added to reduce pH below 4 to stop digestion. Peptides were purified with C18 cartridges (Thermo Pierce™ 89873) and eluted with 50% acetonitrile in LCMS water per manufacturer instructions. Samples

were then dried by speedvac (Thermo Savant), resuspended in 0.1% formic acid in LCMS water, and peptide concentration determined with a quantitative peptide assay per manufacturer instructions (Thermo Pierce™ 23290).

2.2 | LCMS acquisition

All samples were analyzed with a nanoElute® 2 LC which was seamlessly connected online via a column toaster and captive spray (CS) source to an Impact II MS (all Bruker Daltonics). The Impact II operates like a TimsTOF but without the ion mobility dimension [33]. An equal amount of 200 ng of total peptide mix was injected for all samples, except for the BSA standards (12.5 fmol injection). Peptides were separated without trapping on a Pepsep 25 cm $\times$ 150 μm ID $\times$ 1.5 μm particle size column. The optimal gradient was 3%–33% acetonitrile for a duration of 70 min. Other conditions tested included different separation times using the same gradient, different gradients, and different Pepsep columns. Identical acquisition parameters were used in OTOF control 6.2 (Bruker Daltonics) for DDA and DIA except that DDA was operated in MSMS mode (2 s cycle time for selecting MS precursors for MSMS fragmentation, 25 Hz) and DIA was operated in MRM mode (MSMS spectra only, 50 Hz). During MSMS mode, the isolation width for fragmentation was set at 2–3 m/z units, whereas in MRM mode it was 10 ± 0.5 m/z units. For DIA, 75 m/z windows were scanned in each cycle (380–1130 m/z range). Consequently, every scanning cycle required 1.5 s at a frequency of 50 Hz, resulting in the generation of a minimum of 8 data points within a 12 s timeframe. This alignment closely matched the widths of the transition peaks produced under the above-mentioned LC gradient conditions.

2.3 | DDA analysis of reference samples

Data-dependent acquisition (DDA) information was extracted from three samples of each group (C, W, T). This data was employed to capture illustrative MS and MSMS spectra which were subsequently annotated with pertinent details such as peptide sequence, protein accession number, and retention time information, using Fragpipe 19.0 [34]. The reference proteome database for *B. schlosseri*, consisting of 46,519 protein entries was downloaded on May 09, 2023 from https://aniseed.fr/aniseed/download/download_data. The overall proteome database generated for *B. schlosseri* analysis with Fragpipe contained these 46,519 entries plus 118 common contaminants and 46,637 shuffled decoy entries for a total of 93,274 entries. The human RefSeq proteome was retrieved from the NCBI database on April 24, 2023 and was used for conducting a comparative analysis of the HeLa cell extract (Pierce™ # 88328, 200 ng per injection) benchmarking standard. This Refseq proteome contained 175,050 entries plus 118 common contaminants and 175,168 shuffled decoy entries for a total of 350,336 entries. In addition, bovine serum albumin standard (Pierce™ # 88341, 12.5 fmol per injection) was analyzed by DDA. The bovine Refseq proteome contained 64,715 entries plus 118 common contaminants and 64,933 shuffled decoy entries for a total of 129,866 entries and was used to compare bovine BSA peptide intensities, peak widths and retention times before and after acquisition of all *B. schlosseri* samples. To validate consistent instrument performance throughout the analyses we used DataAnalysis 6.2 (Bruker Daltonics). Before subjecting them to Fragpipe analyses, all the raw files were converted to mzML format using MSconvert 3.0. The subsequent Fragpipe analyses were conducted in a closed search mode, using default parameters and the databases outlined above.

2.4 | Generation of a high-quality DIA assay reference library

The DDA data were used to construct a reference DIA assay library, which ensures consistent quantitation of the same set of proteins, peptides, and transitions across all *B. schlosseri* samples. First, annotated spectra information for nine representative samples (three for each group) was generated and stored in pep.xml files using MSFragger 3.7. These files were then imported into Skyline Daily 22.2.1.488 to create a raw spectral library. This raw spectral library was used to generate a DIA assay list for the *B. schlosseri* proteome using the DIA Peptide Search option in Skyline with the following transition settings: precursor charges 1–5, fragment ion charges 1–2, ion types y, b, product ions from ion 3 to last ion, min. m/z 50, max. m/z 2000, ion match tolerance 0.05 m/z, and picking 15 product ions (min. 5). The digestion conditions were set to trypsin, max. 2 missed cleavages, using the *B. schlosseri* proteome as a reference. The Skyline analysis removed 533 redundant proteins from this reference proteome, resulting in 45,986 target entries. Oxidation on Met and carbamidomethylation on Cys were variable structural modifications (max. 3 per peptide). The quantification mode was configured to achieve median equalization

at MS level 2 [35, 36]. The DIA assay list was compiled through the process of creating protein groups, arranging all peptides within their corresponding protein groups, assigning shared peptides exclusively to the protein group with the highest peptide counts, identifying the smallest possible protein list that accounts for all peptides, imposing a peptide length constraint (6–25 amino acids), and introducing an equivalent number of decoys as peptides (45,606, with sequence shuffling). The DIA data from the nine reference samples were initially imported into Skyline before incorporating any additional samples, by using the subsequent full scan settings: Product mass analyzer = centroided; Isolation scheme = results (0.5 margins); Mass accuracy = 15 ppm, Scan collection = only scans within 2 min of the MS/MS identification retention time. The DIA assay library was subsequently enhanced through the process of refining peak outlines with mProphet (evaluating all peak quality criteria and employing decoys for q value scoring). Subsequently, peptides lacking results were eliminated, while the maximum transitions per precursor was capped at 8 (with a preference for larger ions). Additionally, criteria were set requiring a minimum of 5 transitions and a minimal peak quality of 0.5. To prevent bias, no manual adjustments were made to peak boundaries. Nevertheless, peak quality was evaluated using mProphet and a q value cutoff of 0.05 was enforced across all quantitative analyses. In terms of optimizing separation conditions with the nanoElute 2, spectral libraries were constructed using DDA data obtained from three samples of the T population and were then subjected to filtering with encyclopeDIA [37].

2.5 | Relative quantitation of protein abundances

DIA data for all remaining 35 samples were imported into Skyline, utilizing the identical transition settings as above. To enhance iRT alignment, a retention time regression was established using 12 internal standards with the Skyline retention time calculator (referred to as “Bosch”). This calibration was subsequently extended to all peaks across all samples. Next, mProphet peak optimization was performed, using Skyline-generated decoys (matching the count of target peptides). All samples were annotated with two distinct categories (bioreplicate and condition) facilitating the establishment of pairwise statistical comparisons across the three *B. schlosseri* population groups. Comparisons were normalized by equalizing medians at 95% confidence (MS level 2), with proteins serving as the scope and the sum of transitions as the summarization method. A peak detection q value cutoff of 0.05 was enforced for all comparisons. The design sample size module of MSstats 4.2 was used to calculate the fold-change (FC) threshold for significantly different proteins based on the actual data collected for all 44 samples and was set as a statistical power of 0.8 and false discovery rate (FDR) < 0.05 [38]. The data processing module of MSstats was used to generate a QC plot that visualizes protein abundance intensity distributions across all samples. All MSstats group comparisons enforced a q value cutoff of 0.05 to eliminate low-quality peaks from interfering with the quantitation. Pairwise comparisons of proteins were considered statistically significant if they have a FC > 2

and a multiple testing corrected $p < 0.05$. Multiple testing correction was performed with the Benjamini–Hochberg procedure [39].

2.6 | Generation of String reference database and network analysis

FASTA files that were generated with Fragpipe for the nine reference samples (three for each population) were validated and subsequently merged into a unified and definitive FASTA file with PEAKS X Pro (ver. 10.6, BSI Inc., Waterloo, Canada). To comply with the maximum 36 character limit of the public STRING database [40] protein identifiers were abridged and identifier redundancy was removed by deleting the “Boschl.CG.Botznik2013” prefix from all entries. The resultant proteome of *B. schlosseri* encompassing 6612 experimentally validated and unambiguous proteins was then uploaded as a new reference proteome within the STRING database. This proteome was used as the reference for conducting functional enrichment analyses on sets of proteins that display over- and under-representation, as well as those that exhibit the highest and lowest variability within specific *B. schlosseri* populations.

In this experimentally validated reference proteome database, each protein identifier was associated with a corresponding protein name through a Blast2Go batch homology search (Omicsbox 3.0.29, BioBam). STRING network analyses were executed on all protein sets that displayed significant differences between populations, employing the multiple protein search option. The goal was to identify STRING clusters that exhibited notable enrichment or depletion (with FDR < 0.01). These STRING clusters were subsequently consolidated into broader functional categories, streamlining the information and minimizing redundancy to facilitate comprehension. To maintain clarity and eliminate redundancy, other forms of functional enrichments that were obtained from the STRING analysis (GO terms, KEGG pathways, Reactome pathways, Uniprot keywords) were not considered. Nevertheless, it is worth noting that these additional enrichments can be accessed through a subsequent STRING reanalysis of the data included in Table S1. This allows for verification that they mirror the same functional enrichments represented by the STRING clusters.

STRING analyses were conducted on the sets of proteins displaying the highest and lowest variability within each population. These analyses utilized the protein value/rank search option. The values used for these searches corresponded to the degree of abundance fluctuation observed among samples for each protein within a specific population. This variability was calculated by computing the ratio of the standard deviation to the arithmetic mean (Coefficient of variation, CV). Groups of proteins at the upper (>0.8) and lower (<0.2) ends of the CV range were chosen for each population and used for value/rank-based STRING analyses, using the standard gene set STRING analysis workflow. The resultant STRING network clusters were then visualized using default settings, with the exception that network edges were displayed as a single line. The intensity of this line illustrates the confidence (strength) of the connection between nodes. Markov clustering

(MCL) was performed, with the inflation parameter set at 1.5. The resulting network clusters were connected by solid lines while dashed lines indicated nodes belonging to different functional STRING clusters. The network nodes were filled with colors corresponding to the clusters and the intensity of the halos surrounding nodes indicated the protein variation within a given population (color scales ranged from grey to red with a CV of 0.2 to a minimum value, and from gray to blue, with a CV of 0.8 to maximum value).

3 | RESULTS AND DISCUSSION

3.1 | Deep proteome coverage by peptide separation with the nanoElute® 2

We have optimized one of the first commercially available nanoElute® 2 instruments to achieve optimal protein detection in complex mixtures through DDA. The nanoElute 2 is a new nano-LC device designed for ultra-high resolution (UHR) QTOF and TimsTOF mass spectrometers. The capabilities of this new instrument were also harnessed for quantitative label-free DIA proteomics. Optimal peptide separation conditions include the use of a Pepsep 25 cm $\times$ 150 μ m $\times$ 1.5 μ m column, 70 min gradient of 3–33%, 500 nL/min flow rate, 20 μ m zero dead volume CS emitter, and the injection of 200 ng total peptide mixture (Figure S1). Optimal sample preparation was achieved when applying two cycles of 30 s each for 3.0 mm bead homogenization at 3500 rpm, followed by a 4 h trypsin/Lys-C digestion period.

Negligible sample carryover between runs was observed when employing the default column flushing and reconditioning conditions as determined by the Hystar 6 nanoElute® 2 control software. Preliminary trials indicated that best results were obtained by operating the nanoElute® 2 without incorporating a trap column. Instead, C18 cartridges were employed to purify samples prior to injection. This purification process included washes with 5% acetonitrile. However, the starting concentration for the acetonitrile gradient was not increased beyond 3% to minimize the loss of highly hydrophilic peptides and account for distinct peptide binding properties of the C18 matrix in the preinjection cleanup cartridges compared to the separation column.

The workflow detailed above enabled the operation of the nanoElute® 2 in a trap-free mode, effectively minimizing peak width. By implementing these optimized conditions, we identified up to 4930 protein groups, encompassing 28,404 unique peptides in a single sample derived from a single *B. schlosseri* system, even without discerning ion mobility. The average number of protein groups detected in the three populations was 4421 ± 270 (C), 4292 ± 156 (W), and 3853 ± 40 (T), and the corresponding average number of unique peptides was $18,833 \pm 1170$ (C), $17,585 \pm 834$ (W), and $15,572 \pm 358$ (T) (mean $\pm$ SEM, Figure S2, Table S1). These results are in line with the 4041 protein groups and slightly below the count of 26,167 unique peptides identified in a HeLa cell digest [41], a common standard for proteomics experiments (Figure S2). When employing a TimsTOF Pro II mass spectrometer, HeLa cell digest typically yields approximately

10,000 protein group IDs [42], suggesting that an added ion mobility dimension could potentially lead to the identification of nearly 10,000 *B. schlosseri* proteins.

3.2 | A DIA assay library for deep quantitative proteomics of *B. schlosseri*

By consolidating multiple DDA runs to alleviate the impacts of stochastic peak selection and undersampling [43], it becomes possible to increase the detection of peptides with low abundance. This strategy enables the quantification of >7000 *B. schlosseri* protein groups (Figure S3A), collectively represented by almost 40,000 unique peptides (Figure S3B). These data were used to construct a comprehensive *B. schlosseri* DIA assay library consisting of experimentally validated protein sequences. To enhance quantitation robustness, transitions of low quality or with interferences and proteins with fewer than three peptides each (approximately half of the total) were excluded from the final DIA assay library. The remaining 3,426 unambiguous proteins are represented by 22,593 unique peptides and the majority of these peptides (71%) are endowed with 8 quantitative transitions (Figure S3C). Each protein within this DIA assay library is consistently represented by identical peptides and transitions across all samples, ensuring a precise basis for quantitative comparisons [36]. The most abundant transitions in this library are contributed by y_4 – y_8 MSMS ions (Figure S3D).

Of the 3426 proteins included in the *B. schlosseri* DIA assay library, 91 lack close homologs according to the BLAST2GO analysis and are, therefore, unique to *B. schlosseri*. They hold potential for intriguing future functional characterizations (Table S2). Furthermore, among the identified proteins, 368 exhibited corresponding BLAST2GO orthologs designated as “uncharacterized proteins,” reflecting the limited understanding of their functions. Large multiprotein complexes are also well-represented in the DIA assay library, including the ubiquitination machinery (31 subunits), 26S proteasome (24 subunits), cytosolic ribosome (44 subunits), mitochondrial respiratory chain complex (30 subunits), nuclear pore complex (12 subunits), nucleolus complex (10 subunits), cilia- and flagella-associated protein complex (23 subunits), mRNA spliceosome (37 subunits), sorting nexin complex (10 subunits), T-complex (13 subunits), and THO-complex (4 subunits). These large protein complexes reside in different cell compartments including the cytosol, nucleus, mitochondria, and various plasma membrane and endosomal compartments [44–49], confirming that each of these cellular compartments is well represented in the DIA assay library.

Across all samples, the MSMS transitions had a mass accuracy of <15 ppm (<10 ppm for the vast majority, Figure S3E). The quality of recorded peaks from all samples was high, signified by mProphet q values < 0.05 (Figure S3F). While some peaks slightly deviated from the predicted retention time, their frequency was minimal, and thus no manual adjustments to peak boundaries were made to ensure consistency and eliminate potential bias (Figure S3G). To establish a pertinent fold change (FC) threshold for statistical significance, a

power analysis encompassing the entire dataset was performed with MSstats [38]. This analysis indicated a threshold FC value of 2-fold, relevant for a sample size of at least 12 replicates per group (Figure S3H). Furthermore, the MSstats power analysis also indicated that the distribution of relative intensity for peptides representing each protein remained comparable among all samples (Figure S3I). Collectively, these data illustrate the high quality of the *B. schlosseri* DIA assay library.

3.3 | Proteome differences between *B. schlosseri* laboratory and field populations

Quantitative assessment of relative protein abundances in the Santa Barbara samples (lab C vs. field W) unveiled 100 proteins that were significantly upregulated (55 with STRING AC) and 237 that were significantly downregulated (121 with STRING AC) in laboratory-raised versus field populations (Figure 2A). A substantial portion of these differentially regulated proteins clustered within the extreme 5% of the highest and lowest percentiles STRING input value rank distribution (Figure 2B). Many of the up- and down-regulated proteins that lack STRING ACs, have predicted functions that align with the functional categories represented by the proteins having a STRING AC. Particularly noteworthy are cytoskeleton, ribosomal translation, and extracellular matrix (ECM)/immunity functional domains (see below, Table S2). These functions contribute to dynamic remodeling of vascular tissue [50], whole body regeneration [51], and allogeneic immunity [52] in *B. schlosseri*, suggesting that laboratory conditions significantly impact these processes. However, among the significantly upregulated and downregulated proteins, 15 and 51, (including the two most strongly) respectively, did not exhibit sequences similar to any known protein. The experimental characterization of these proteins holds substantial potential for revealing novel domains and functions that confer adaptive advantages in the face of intense selection pressure (e.g., laboratory environments).

Both sets of significant proteins having annotated STRING ACs (up- and down-regulated) display a pronounced overlap in functional categories. This observation implies a comprehensive disparity in the corresponding functions, achieved through the preferential involvement of distinct proteins governing these functions in laboratory-bred versus field populations. These functions encompass glycolipid and lipid metabolism (4/17), cytoskeleton (axonemal cilia, actin, microtubules, 5/19), xenobiotic metabolism (12/15), chromatin and DNA repair/replication (2/9), ion transport (2/5), intracellular signaling (4/7), RNA metabolism and processing (6/9), glycoprotein and sugar metabolism (3/9), ECM and immunity (6/31), energy metabolism (5/12), amino acid metabolism (9/12), ribosomal translation (4/4), endoplasmic reticulum (ER) processing and vesicle transport (9/10), and proteolysis (6/31) (corresponding up/down-regulated protein numbers are in parentheses, Figure 2C,D).

To determine whether these functional differences persist for other field versus laboratory *B. schlosseri* comparisons, we compared the proteome of the Santa Barbara laboratory-raised population (C) with the

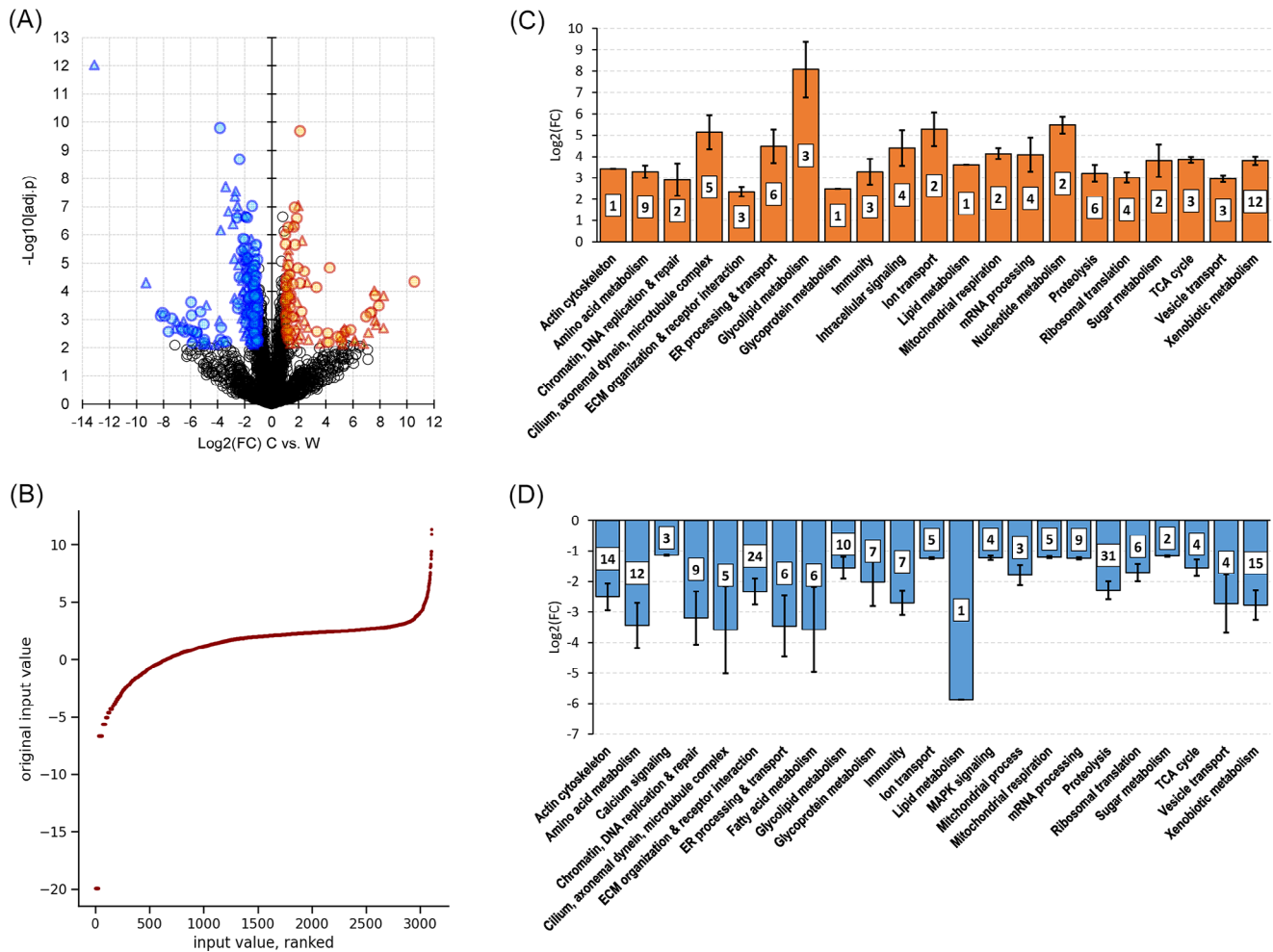


FIGURE 2 Relative protein abundances in populations C versus W. (A) Volcano plot visualizing statistically significant ($\text{FC} > 2$ and corrected $p < 0.05$) proteins that are more (orange) and less (blue) abundant in C than W. Colored circles represent proteins with a STRING annotation while colored triangles are proteins not included in the STRING database (i.e., proteins without a sufficiently similar ortholog). (B) STRING input value rank distribution based on $\log_2(\text{FC})$ when comparing C versus W. STRING cluster categories for proteins that are significantly more (C) and less (D) abundant in C versus W. Data represent mean $\pm$ SEM for $\log_2(\text{FC})$ of proteins in each category and numbers inside bars indicate the corresponding numbers of proteins. Some significant proteins were mapped to multiple functions.

remote Des Moines, WA field population (T), which is located in a different habitat. Relative protein abundance quantitation revealed 78 proteins (40 with STRING ACs) that are significantly upregulated and 647 proteins (479 with STRING AC) that are significantly downregulated in the “C” versus the “T” *B. schlosseri* populations (Figure 3A). Similar to previous comparisons, the majority of these significantly different proteins clustered within the extreme 5% of highest and lowest percentiles of the STRING input value rank distribution (Figure 3B). The functions of most proteins that were up- and down-regulated in the lab population “C” but lacked STRING ACs aligned with the functional categories represented by proteins with STRING ACs (see below, Table S2). Among these proteins, the 15 significantly upregulated and 55 significantly downregulated proteins lacking STRING ACs, encompassed those that were most prominently regulated in either direction. These proteins represent promising candidates for uncovering novel domains and functions associated with the effects of artificial selection under laboratory conditions [53].

Despite the considerable disparities in habitat and latitude between the field populations (Santa Barbara, CA and Des Moines, WA), similar functional variations were observed in both field versus laboratory-raised population comparisons (Des Moines vs. laboratory and Santa Barbara vs. laboratory). In addition, in both comparisons, more proteins were significantly down- than up-regulated in the laboratory-raised population for all functional categories. The proteins that displayed significant regulation in the laboratory versus field populations were implicated in a range of functional categories, including glycolipid and lipid metabolism (2/42), cytoskeleton (axonemal cilia, actin, microtubules, 8/75), xenobiotic metabolism (8/15), chromatin and DNA repair/replication (5/40), ion transport (2/17), intracellular signaling (0/72), RNA metabolism and processing (8/74), glycoprotein and sugar metabolism (0/51), extracellular matrix and immunity (8/52), energy metabolism (2/55), amino acid metabolism (7/46), ribosomal translation (0/47), ER processing and vesicle transport (11/91), and proteolysis (6/61) (up-/down-regulated protein numbers in parentheses,

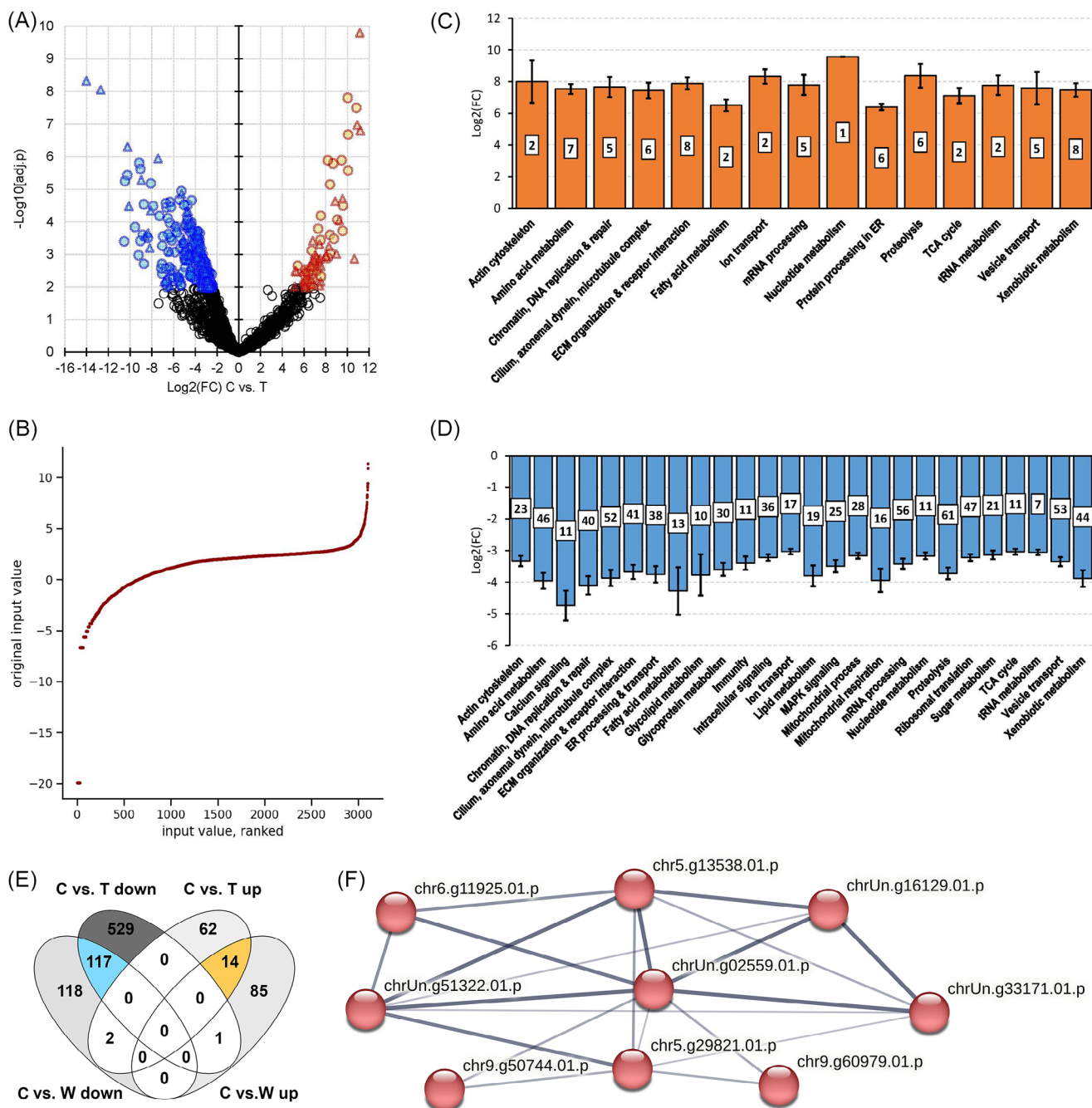


FIGURE 3 Relative protein abundances in populations C versus T. (A) Volcano plot visualizing statistically significant ($\text{FC} > 2$ and corrected $p < 0.05$) proteins that are more (orange) and less (blue) abundant in C than T. Colored circles represent proteins with a STRING annotation while colored triangles are proteins not included in the STRING database (i.e., proteins without a sufficiently similar ortholog). (B) STRING input value rank distribution based on Log2(FC) when comparing C versus T. STRING cluster categories for proteins that are significantly more (C) and less (D) abundant in C versus T. (E) Venn diagram of common proteins that are significantly up-regulated (orange shading) and down-regulated (blue shading) in the laboratory population (C) compared to both field populations (W and T). (F) STRING MCL network showing 9 proteins that are consistently down-regulated in the laboratory population and belong to a functional cluster of RNA processing proteins. Data represent mean $\pm$ SEM for Log2(FC) of proteins in each category and numbers inside bars indicate the corresponding numbers of proteins. Some significant proteins were mapped to multiple functions.

Figure 3C,D). As expected, the proteome differences between the laboratory-raised and Des Moines field populations were more pronounced compared to those between the laboratory-raised and Santa Barbara populations since the laboratory-raised population originated from the Santa Barbara field population.

Almost half (117 out of 235) of the proteins that were significantly lower and 14% (14 out of 100) that were significantly higher in abundance in the lab population (C) compared to the Santa Barbara field population (W) were also significantly different in the same direction when comparing the lab population to the Des Moines field population (T) (Figure 2E). The 14 proteins that are consistently upregulated in the lab population relative to both field populations include multiple proteins involved in translation and cytoskeleton organization, however, they do also include 3 proteins without STRING AC and do not represent an enriched STRING cluster (Table S4). The 117 proteins that are consistently down-regulated in the lab population relative to both field populations include 12 proteins without STRING AC and are not significantly enriched for any functional STRING cluster (Table S4). Nevertheless, MCL clustering of these 117 proteins reveals that 9 of them belong to a functional cluster of RNA processing proteins (Figure 3F). This observation suggests that specific aspects of RNA processing are down-regulated when transferring animals from the field into a laboratory culture setting.

The observed changes in fatty acid metabolism between the lab and field populations (Figures 2 and 3) may also be part of the changes in the appearance or changes in the metabolism of Thraustochytrids, widely distributed marine heterotrophic microorganisms known for their potential to accumulate docosahexaenoic acid (DHA)-enriched lipids [54]. Lipid metabolism in these associated saprophytic organisms is directly affected by the environmental conditions (including pH, temperature, oxygen concentration, salinity, carbon and nitrogen sources) [55]. As Thraustochytrids are very common in *B. schlosseri* colonies collected either from various field locations or from laboratory bred colonies [56, 57], the up-/down-regulated glycolipid and lipid metabolism results, may also be the outcomes of changes in the Thraustochytrids' metabolism [58].

The overall difference between the proteomes of the two field populations is notably smaller compared to the differences observed between the laboratory-raised and either one of the field populations (Figure 4A,B). Out of the 46 proteins that exhibited significantly higher abundance in the Des Moines population than in the Santa Barbara field population, only 18 could be mapped onto STRING orthologs. Likewise, only 11 of the 30 proteins that had significantly lower abundance in the Des Moines population had a STRING AC. The remaining proteins lacked sufficient homology to warrant assignment of a functional STRING ortholog. These proteins, which potentially hold novel functions, encompass three putative isoforms of 4-coumarate-CoA ligase 1, standing out as the most prominently elevated proteins in Des Moines over Santa Barbara field populations. Additionally, two unknown proteins were among the most substantially reduced proteins in the Des Moines versus Santa Barbara field population comparison (Figure 4A). The presence of 4-coumarate-CoA ligases is a characteristic feature of microbes and plants, playing roles in sec-

ondary metabolite and lignin metabolism as well as cell signaling within those taxonomic groups [59, 60]. Consequently, the identification of these enzymes within the *B. schlosseri* proteome raises the possibility of common commensals or parasites, being included in the original reference genome. The significantly lower abundance of these proteins in the Des Moines population may indicate that such symbionts or parasites are less abundant than in the Santa Barbara field population. A recent study on another botryllid ascidian has revealed that under stress conditions that lead to a torpor state, a novel lineage of the commensal bacterium *Endozoicomonas*, was dominant in torpor animals, and potentially occupied hemocyte types that were only detected in torpid animals [61]. This bacterial symbiont is found in close association with diverse marine invertebrate hosts, including reef-building corals, sponges, bryozoans, sea slugs, and other tunicates [62–66]. It is also plausible that these proteins are indeed genuine *B. schlosseri* proteins, potentially serving functions beyond the scope of 4-coumarate-CoA ligase activity, which remain undiscovered.

Several protein functional categories were significantly different between the Santa Barbara laboratory-raised population relative to either one of the two field populations (chromatin and DNA repair/replication, ion transport, intracellular signaling) but not between the two disparate field populations. Furthermore, even those functional categories that were significantly different between both field populations included notably fewer proteins than for the laboratory population. Significantly up-/down-regulated proteins in the Santa Barbara compared to Des Moines field populations contribute to glycolipid and lipid metabolism (0/2), cytoskeleton (axonemal cilia, actin, microtubules, 2/4), xenobiotic metabolism (4/1), energy metabolism (2/0), amino acid metabolism (8/0), RNA metabolism and processing (1/0), glycoprotein and sugar metabolism (0/2), extracellular matrix and immunity (1/5), ribosomal translation (1/0), ER processing and vesicle transport (2/3), and proteolysis (2/3) (up-/down-regulated protein numbers are shown in parentheses, Figure 4C,D). Collectively, these results suggest that extended laboratory rearing across multiple generations exerts a more pronounced impact on the proteome of *B. schlosseri* compared to the variations stemming from the diversity in natural habitats. It is important to note that this conclusion is based on the comparison of the laboratory-raised population against only two field sites; however, these sites differ significantly in terms of latitude (1800 km apart) and environmental parameters such as temperature and photoperiod. Similar substantial differences in proteomes between laboratory and field populations have been previously documented in bacteria [67–69], suggesting that substantial proteome changes are a common phenomenon across taxa when adapting organisms to laboratory rearing conditions.

3.4 | *B. schlosseri* proteins exhibiting the lowest and highest coefficient of variation between individuals

To identify *B. schlosseri* proteins demonstrating the lowest and highest CV within each population, we assessed individual-to-individual variance for each protein within the three populations. All three

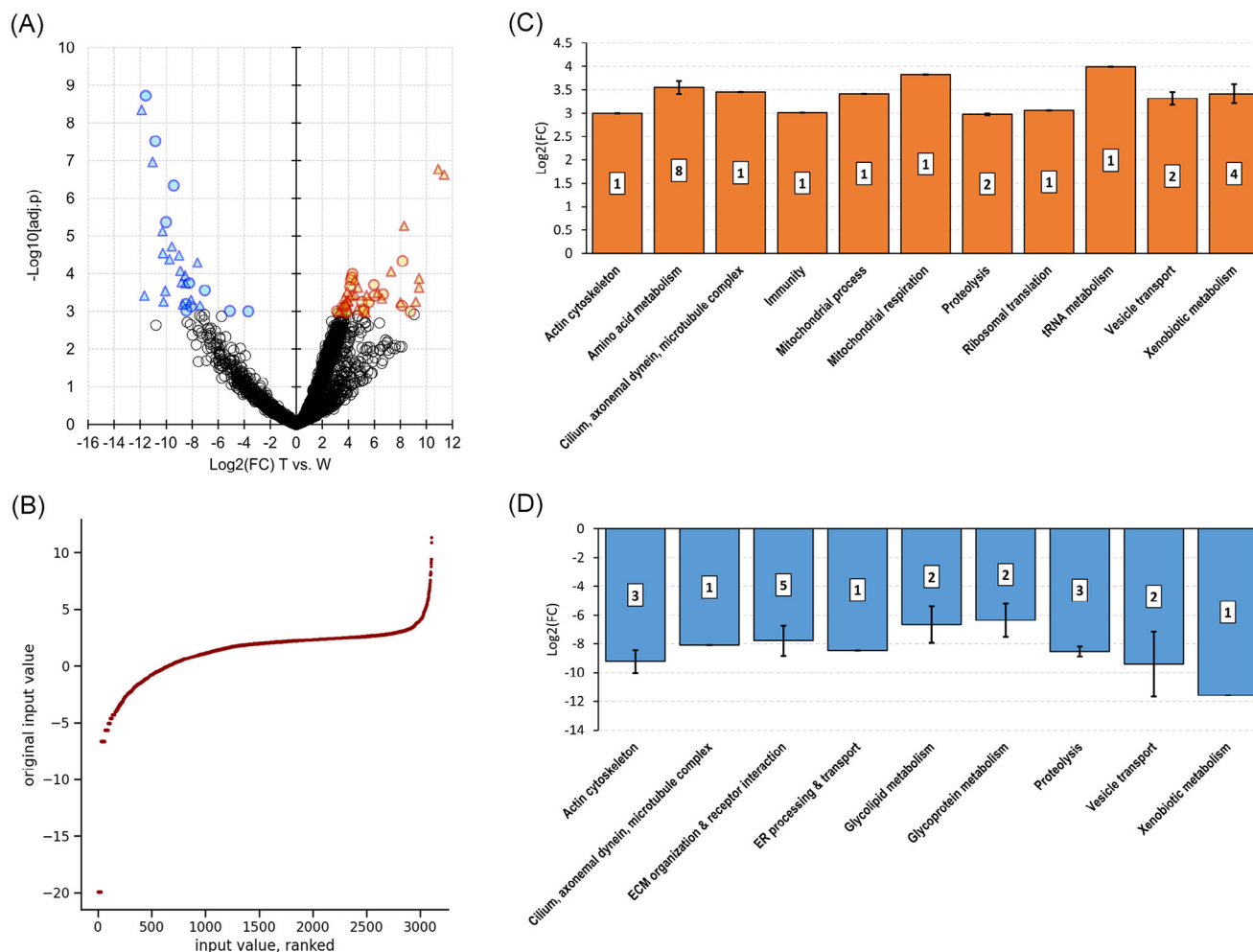


FIGURE 4 Relative protein abundances in populations T versus W. (A) Volcano plot visualizing statistically significant ($FC > 2$ and corrected $p < 0.05$) proteins that are more (orange) and less (blue) abundant in T than W. Colored circles represent proteins with a STRING annotation while colored triangles are proteins not included in the STRING database (i.e., proteins without a sufficiently similar ortholog). (B) STRING input value rank distribution based on $\text{Log}_2(FC)$ when comparing T versus W. STRING cluster categories for proteins that are significantly more (C) and less (D) abundant in T versus W. Data represent mean $\pm$ SEM for $\text{Log}_2(FC)$ of proteins in each category and numbers inside bars indicate the corresponding numbers of proteins. Some significant proteins were mapped to multiple functions.

populations exhibited similar STRING variance rank distributions (Figure 5A–C). The majority of proteins are tightly regulated across all populations with 93% (C and W)—96% (T) having a $CV < 0.8$ and 75% (C and W)—86% (T) having a $CV < 0.5$. However, some proteins exhibited unusually low ($CV < 0.2$) or high ($CV > 0.8$) variances, quantified in terms of the standard deviation (STD) of abundance divided by the mean abundance. Elevated interindividual variance in protein abundance denotes fewer functional constraints while low interindividual variance suggests robust functional constraints [70, 71]. Sets of proteins displaying tight co-variations in abundance typically signal their involvement in the same biological process or function [72]. Likewise, proteins with similarly extreme CV (either very low or very high) may be indicative of modestly or highly constraint functions. Consequently, we examined whether comparable functions are manifested within the high- and low-variance protein sets across all three populations. This analysis revealed that a majority of the most and the least

variable proteins were population-specific (Figure 5D,E). This result may potentially mirror significant disparity (in the most variable) and concordance (in the least variable) of allele frequencies at corresponding genetic loci. This has been suggested for other organisms such as mice [73] and fruit flies [74]. It should be pointed out that adaptation often occurs as a result of sequence variation in cis-regulatory elements rather than coding sequences [75]. The highly variable proteins exclusive to the laboratory-raised population of *B. schlosseri* might potentially stem from genomic loci characterized by heightened allelic (including corresponding cis-regulatory regions) diversity and distinct adaptive significance during artificial selection for laboratory rearing conditions.

A total of 33 proteins consistently exhibited remarkably low interindividual CV in all three populations (Figure 5D). Many of these co-varying and overrepresented proteins cluster within a single STRING functional network, consisting of spliceosomal mRNA

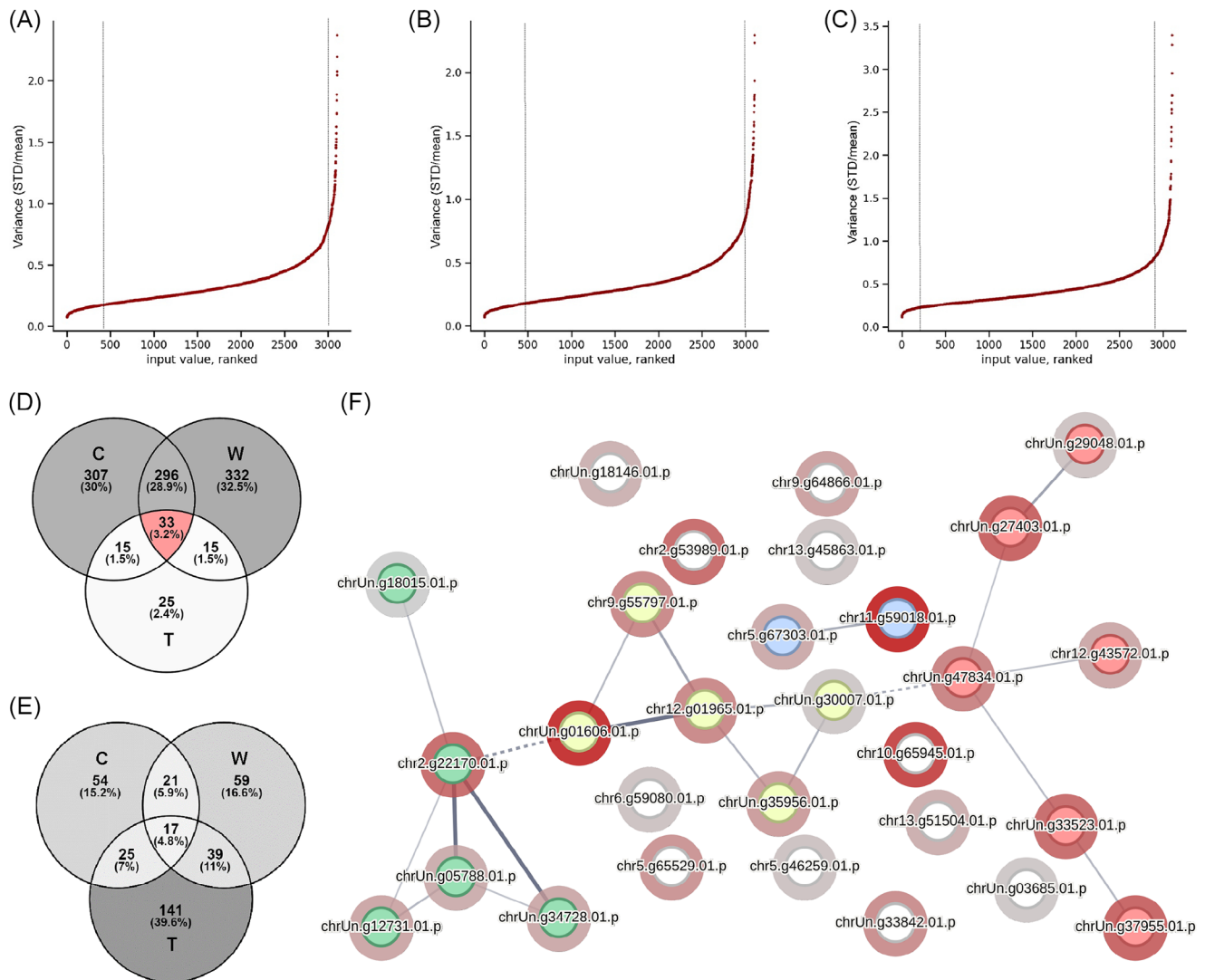


FIGURE 5 Most and least variable proteins within a single *B. schlosseri* population for all three populations. STRING rank distribution of protein variance (standard deviation divided by mean, STD/mean) is depicted for groups C (A), W (B), and T (C). Venn diagrams illustrating protein overlap between all three groups are shown for the least (D, STD/mean < 0.2) and most (E, STD/mean > 0.8) variable proteins. (F) STRING network containing all 33 proteins with consistently low individual variability (STD/mean < 0.2) across all three populations. MCL clusters are connected by edges (line thickness = strength of functional relatedness) and indicate that spliceosome (STRING:CL263, red filling) and Wnt signaling (KEGG map04310, green filling) are overrepresented in the least variable (most tightly regulated) proteins. The redness of the halo surrounding nodes indicates individual variability across all three populations (STD/mean abundance ranging from 0.14, red, to 0.19, gray).

processing proteins and Wnt signaling proteins. Wnt signaling holds pivotal significance in developmental regulation and stemness [76], making it intriguing that proteins within this pathway manifest the lowest interindividual CV across all populations studied. Although spliceosomal mRNA processing represents one of the functions differing significantly between the three populations, the interindividual variability of spliceosomal proteins within any given population is minimal across all populations. This result implies that, although spliceosomal functions are tightly controlled, they are regulated at different overall levels in different populations [77].

The set of 17 proteins exhibiting high interindividual abundance variance across all populations did not align with any STRING network

(Figure 5E). However, a significant portion of these proteins are characterized by their representation of ECM proteases and other matrix proteins along with their receptors. The high CV observed in the abundance of these proteins implies that individual *B. schlosseri* colonies originating from the same population demonstrate the most divergence in terms of their extracellular matrix composition (Table S3). For the field populations this outcome might be anticipated, given the wide variability in substrate to which *B. schlosseri* colonies attach in field settings [78, 79]. The substrate utilized for laboratory cultures is highly uniform (usually glass slides) but very different from natural substrates [17]. Therefore, different attachment substrates may contribute to the considerable interindividual variance in ECM protein

abundance of *B. schlosseri*. Moreover, ECM architectures, molecular compositions, and dynamic expression profiles of ECM proteins in tunicates depend on developmental stage, including blastogenesis [80, 81], and are altered during inflammatory reactions and noncompatible allorecognition [82, 83].

3.5 | Concluding remarks

This study represents the first deep proteome analysis of any colonial tunicate. It reveals quantitative alterations in the *B. schlosseri* proteome that become evident through comparisons of distinct populations (laboratory reared and field collected) originating from the same founder population. Significant proteome variation was observed when organisms are shifted from their natural habitat (in situ) to controlled husbandry settings (ex situ). The extent of proteome variation between two genetically closely related populations (field vs. laboratory raised Santa Barbara colonies) likely signifies shifts occurring as *B. schlosseri* cells transition from their natural in vivo environment to artificial in vitro culture conditions. This observation is critical for future studies focusing on the development of cell cultures derived from laboratory-bred *B. schlosseri* colonies [84, 85]. This study tackles the first steps toward addressing a knowledge gap that is of broad interest to cell biologists, that is uncovering the mechanisms that govern cell proliferation and cellular senescence, as well as the processes through which cells achieve immortality in vitro. This lack of knowledge is particularly significant in aquatic invertebrates (including *B. schlosseri*), as there is currently only a single report of successful in vitro cell immortalization for any of them [86]. Our results indicate that many unknown proteins lacking a STRING functional annotation hold significance for processes associated with the regulation of proliferation during in situ to ex situ transition of *B. schlosseri*. Consequently, they present captivating prospects for future endeavors to unravel their functions.

ACKNOWLEDGMENTS

We thank Ayelet Voskoboynik (Stanford Univ., Hopkins Marine Station) for her advice on *B. schlosseri* biology, culture, and genomic resources. We are also grateful to Brian Searle (Ohio State Univ.) and the Skyline and Panorama developer team (Univ. Washington, Seattle) for support with Skyline, Panorama, and encyclopeDIA. Gina Jones and Lindsay Overstreet (Univ. Washington Tacoma) are thanked for assisting with field collections. Finally, we are grateful to the staff of Des Moines and Santa Barbara marinas for providing access to field sites. This research was funded by NSF grants MCB-2127516 and MCB-2127517 (D.K. and, A.M.G., respectively) and BSF grant 2021650 (B.R.).

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

All raw LCMS data were deposited in PanoramaPublic and can be accessed at <https://panoramaweb.org/bosch01kl.url>, which is also the

access point to the DIA assay library and Skyline analyses of DIA data for all samples. In addition, all raw data and DIA data were deposited at ProteomeXchange and are publicly accessible under AC PXD043046. The public doi address for these data is <https://doi.org/10.6069/5fa9-3r35>. The experimentally validated *B. schlosseri* STRING reference proteome can be accessed at <https://version-11-5.string-db.org/organism/STRG0025DCR>.

ORCID

Dietmar Kltz  <https://orcid.org/0000-0002-0745-9645>

REFERENCES

- Manni, L., Anselmi, C., Cima, F., Gasparini, F., Voskoboynik, A., Martini, M., Peronato, A., Burighel, P., Zaniolo, G., & Ballarin, L. (2019). Sixty years of experimental studies on the blastogenesis of the colonial tunicate *Botryllus schlosseri*. *Developmental Biology*, 448(2), 293–308. <https://doi.org/10.1016/j.ydbio.2018.09.009>
- Berrill, N. J. (1950). *The Tunicata: With an account of the British species* (pp. 354). The Ray Society.
- Reem, E., Douek, J., Paz, G., Katzir, G., & Rinkevich, B. (2017). Phylogenetics, biogeography and population genetics of the ascidian *Botryllus schlosseri* in the Mediterranean Sea and beyond. *Molecular Phylogenetics and Evolution*, 107, 221–231. <https://doi.org/10.1016/j.ympev.2016.10.005>
- Zwahlen, J., Reem, E., Douek, J., & Rinkevich, B. (2022). Long-term population genetic dynamics of the invasive ascidian *Botryllus schlosseri*, lately introduced to Puget Sound (Washington, USA) marinas. *Estuarine, Coastal and Shelf Science*, 270, 107840. <https://doi.org/10.1016/j.ecss.2022.107840>
- Ben-Hamo, O., & Rinkevich, B. (2021). *Botryllus schlosseri*: A model colonial species in basic and applied studies. In A. Boutet, & B. Schierwater (Eds.), *Handbook of marine model organisms in experimental biology* (1st ed., pp. 385–402). CRC Press.
- Manni, L., Zaniolo, G., Cima, F., Burighel, P., & Ballarin, L. (2007). *Botryllus schlosseri*: A model ascidian for the study of asexual reproduction. *Developmental Dynamics*, 236(2), 335–352. <https://doi.org/10.1002/dvdy.21037>
- Rosengarten, R. D., & Nicotra, M. L. (2011). Model systems of invertebrate allorecognition. *Current Biology*, 21(2), R82–92. <https://doi.org/10.1016/j.cub.2010.11.061>
- Reem, E., Douek, J., Katzir, G., & Rinkevich, B. (2013). Long-term population genetic structure of an invasive urochordate: The ascidian *Botryllus schlosseri*. *Biological Invasions*, 15, 225–241. <https://doi.org/10.1007/s10530-012-0281-2>
- Stoner, D., Ben-Shlomo, R., & Weissman, I. (2002). Genetic variability of *Botryllus schlosseri* invasions to the east and west coasts of the USA. *Marine Ecology Progress Series*, 243, 93–100. <https://doi.org/10.3354/meps243093>
- Karahan, A., Douek, J., Paz, G., & Rinkevich, B. (2016). Population genetics features for persistent, but transient, *Botryllus schlosseri* (Urochordata) congregations in a central Californian marina. *Molecular Phylogenetics and Evolution*, 101, 19–31. <https://doi.org/10.1016/j.ympev.2016.05.005>
- Hirose, E., Shirae, M., & Saito, Y. (2003). Ultrastructures and classification of circulating hemocytes in 9 botryllid ascidians (chordata: Ascidacea). *Zoological Science*, 20(5), 647–656. <https://doi.org/10.2108/zsj.20.647>
- Qarri, A., Kltz, D., Gardell, A., & Rinkevich, Y. (2023). Improved media formulations for primary cell cultures derived from a colonial urochordate. *Cells*, 12, 1709. <https://doi.org/10.3390/cells12131709>
- Gasparini, F., Manni, L., Cima, F., Zaniolo, G., Burighel, P., Caicci, F., Franchi, N., Schiavon, F., Rigon, F., Campagna, D., & Ballarin, L. (2015).

- Sexual and asexual reproduction in the colonial ascidian *Botryllus schlosseri*. *Genesis* (New York, N.Y.: 2000), 53(1), 105–120. <https://doi.org/10.1002/dvg.22802>
14. Milkman, R. (1967). Genetic and developmental studies on *Botryllus schlosseri*. *Biological Bulletin*, 132(2), 229–243. <https://doi.org/10.2307/1539891>
15. Phillippi, A., & Yund, P. (2017). Self-fertilization and inbreeding depression in three ascidian species that differ in genetic dispersal potential. *Marine Biology*, 164, 179. <https://doi.org/10.1007/s00227-017-3214-x>
16. Brunetti, R., Beghi, L., Bressan, M., & Marin, M. (1980). Combined effects of temperature and salinity on colonies of *Botryllus schlosseri* and *Botrylloides leachi* (Ascidacea) from the Venetian Lagoon. *Marine Ecology Progress Series*, 2, 303–314. <https://doi.org/10.3354/meps002303>
17. Rinkevich, B., & Shapira, M. (1998). An improved diet for inland broodstock and the establishment of an inbred line from *Botryllus schlosseri*, a colonial sea squirt (Ascidacea). *Aquatic Living Resources*, 11(3), 163–171. [https://doi.org/10.1016/S0990-7440\(98\)80113-7](https://doi.org/10.1016/S0990-7440(98)80113-7)
18. Boyd, H. C., Brown, S. K., Harp, J. A., & Weissman, I. L. (2016). Growth and sexual maturation of laboratory cultured Monterey *Botryllus schlosseri*. *Biological Bulletin*, 170, 91–109. <https://doi.org/10.2307/1541383>
19. Rinkevich, Y., Paz, G., Rinkevich, B., & Reshef, R. (2007). Systemic bud induction and retinoic acid signaling underlie whole body regeneration in the urochordate *Botrylloides leachii*. *PLoS Biology*, 5(4), e71. <https://doi.org/10.1371/journal.pbio.0050071>
20. Nourizadeh, S., Kassmer, S., Rodriguez, D., Hiebert, L. S., & De Tomaso, A. W. (2021). Whole body regeneration and developmental competition in two botryllid ascidians. *EvoDevo*, 12(1), 15. <https://doi.org/10.1186/s13227-021-00185-y>
21. Rodriguez, D., Kassmer, S. H., & De Tomaso, A. W. (2017). Gonad development and hermaphroditism in the ascidian *Botryllus schlosseri*. *Molecular Reproduction and Development*, 84(2), 158–170. <https://doi.org/10.1002/mrd.22661>
22. Mojica, E. A., & Kültz, D. (2022). Physiological mechanisms of stress-induced evolution. *Shi Yan Sheng Wu Xue Bao = Journal of Experimental Biology*, 225(1), jeb243264. <https://doi.org/10.1242/jeb.243264>
23. Wessel, G. M., Morita, S., & Oulhen, N. (2020). Somatic cell conversion to a germ cell lineage: A violation or a revelation? *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 22952, 1–14. <https://doi.org/10.1002/jez.b.22952>
24. Heng, H. H. (2019). *Genome chaos: Rethinking genetics, evolution, and molecular medicine* (1st ed., p. 556). Academic Press.
25. Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., & Campbell, P. J. (2017). Universal patterns of selection in cancer and somatic tissues. *Cell*, 171(5), 1029–1041e21. <https://doi.org/10.1016/j.cell.2017.09.042>
26. Vendramin, R., Litchfield, K., & Swanton, C. (2021). Cancer evolution: Darwin and beyond. *Embo Journal*, 40(18), e108389. <https://doi.org/10.15252/embj.2021108389>
27. Rotter, A., Barbier, M., Bertoni, F., Bones, A., Cancela, L., Carlsson, J., Carvalho, M., Ceglowska, M., Chirivella-Martorell, J., Dalay, M., Cueto, M., Dailianis, T., Deniz, I., Diaz-Marrero, A., Drakulovic, D., Dubnika, A., Edwards, C., Einarsson, H., Erdogan, A., ... Vasquez, M. (2021). The essentials of marine biotechnology. *Frontiers in Marine Science*, 8(8), 629629. <https://doi.org/10.3389/fmars.2021.629629>
28. Kassmer, S. H., Nourizadeh, S., & De Tomaso, A. W. (2019). Cellular and molecular mechanisms of regeneration in colonial and solitary Ascidians. *Developmental Biology*, 448(2), 271–278. <https://doi.org/10.1016/j.ydbio.2018.11.021>
29. Taketa, D. A., & De Tomaso, A. W. (2015). *Botryllus schlosseri* Allorecognition: Tackling the enigma. *Developmental and Comparative Immunology*, 48(1), 254–265. <https://doi.org/10.1016/j.dci.2014.03.014>
30. Delsuc, F., Brinkmann, H., Chourrout, D., & Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature*, 439(7079), Article 7079. <https://doi.org/10.1038/nature04336>
31. Rabinowitz, C., Alfassi, G., & Rinkevich, B. (2009). Further portrayal of epithelial monolayers emergent de novo from extirpated ascidians palaeal buds. *In Vitro Cellular & Developmental Biology - Animal*, 45(7), 334–342. <https://doi.org/10.1007/s11626-009-9179-4>
32. Qarri, A., Rinkevich, Y., & Rinkevich, B. (2022). Improving the yields of blood cell extractions from *Botryllus schlosseri* vasculature. In *Advances in aquatic invertebrate stem cell research* (pp. 335–349). MDPI. <https://doi.org/10.3390/books978-3-0365-1635-6-10>
33. Beck, S., Michalski, A., Raether, O., Lubeck, M., Kaspar, S., Goedecke, N., Baessmann, C., Hornburg, D., Meier, F., Paron, I., Kulak, N. A., Cox, J., & Mann, M. (2015). The Impact II, a very high-resolution quadrupole time-of-flight instrument (QTOF) for deep shotgun proteomics. *Molecular & Cellular Proteomics*, 14(7), 2014–2029. <https://doi.org/10.1074/mcp.M114.047407>
34. Demichev, V., Szyrwiell, L., Yu, F., Teo, G. C., Rosenberger, G., Niewianda, A., Ludwig, D., Decker, J., Kaspar-Schoenefeld, S., Lilley, K. S., Müllender, M., Nesvizhskii, A. I., & Ralser, M. (2022). Dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. *Nature Communications*, 13(1), 3944. <https://doi.org/10.1038/s41467-022-31492-0>
35. Pino, L. K., Searle, B. C., Bollinger, J. G., Nunn, B., MacLean, B., & MacCoss, M. J. (2017). The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrometry Reviews*, 39(3), 229–244. <https://doi.org/10.1002/mas.21540>
36. Root, L., Campo, A., MacNiven, L., Cnaani, A., & Kültz, D. (2021). A data-independent acquisition (DIA) assay library for quantitation of environmental effects on the kidney proteome of *Oreochromis niloticus*. *Molecular Ecology*, 21, 2486–2503. <https://doi.org/10.1111/1755-0998.13445>
37. Searle, B. C., Swearingen, K. E., Barnes, C. A., Schmidt, T., Gessulat, S., Küster, B., & Wilhelm, M. (2020). Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nature Communications*, 11(1), 1548. <https://doi.org/10.1038/s41467-020-15346-1>
38. Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., & Vitek, O. (2014). MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics*, 30(17), 2524–2526. <https://doi.org/10.1093/bioinformatics/btu305>
39. Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
40. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & von Mering, C. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>
41. Mische, S. M., Cole, R., Mason, C. E., Weintraub, S. E., & White, A. N. (2020). The ABRF 2020 meeting: Empowering team science. *Journal of Biomolecular Techniques*, 31(Suppl), 01–11. https://doi.org/10.7171/jbt.20-31_Suppl.001
42. Meier, F., Brunner, A.-D., Frank, M., Ha, A., Bludau, I., Voytik, E., Kaspar-Schoenefeld, S., Lubeck, M., Raether, O., Bache, N., Aebbersold, R., Collins, B. C., Röst, H. L., & Mann, M. (2020). diaPASEF: Parallel accumulation–serial fragmentation combined with data-independent acquisition. *Nature Methods*, 17(12), Article 12. <https://doi.org/10.1038/s41592-020-00998-0>

43. Guo, T., & Aebersold, R. (2023). Recent advances of data-independent acquisition mass spectrometry-based proteomics. *Proteomics*, 23(7–8), 2200011. <https://doi.org/10.1002/pmic.202200011>
44. Schipper-Krom, S., Sanz, A. S., van Bodegraven, E. J., Speijer, D., Florea, B. I., Ova, H., & Reits, E. A. (2019). Visualizing proteasome activity and intracellular localization using fluorescent proteins and activity-based probes. *Frontiers in Molecular Biosciences*, 6, 56.
45. Wójcik, C., & DeMartino, G. N. (2003). Intracellular localization of proteasomes. *International Journal of Biochemistry & Cell Biology*, 35(5), 579–589. [https://doi.org/10.1016/s1357-2725\(02\)00380-1](https://doi.org/10.1016/s1357-2725(02)00380-1)
46. Jimeno, S., & Aguilera, A. (2010). The THO complex as a key mRNP biogenesis factor in development and cell differentiation. *Journal of Biology*, 9(1), 6. <https://doi.org/10.1186/jbiol217>
47. Chen, X.-Q., Fang, F., Florio, J. B., Rockenstein, E., Masliah, E., Mobley, W. C., Rissman, R. A., & Wu, C. (2018). T-complex protein 1-ring complex enhances retrograde axonal transport by modulating tau phosphorylation. *Traffic (Copenhagen, Denmark)*, 19(11), 840–853. <https://doi.org/10.1111/tra.12610>
48. Pessa, H. K. J., Will, C. L., Meng, X., Schneider, C., Watkins, N. J., Perälä, N., Nymark, M., Turunen, J. J., Lüthmann, R., & Frilander, M. J. (2008). Minor spliceosome components are predominantly localized in the nucleus. *Proceedings National Academy of Science USA*, 105(25), 8655–8660. <https://doi.org/10.1073/pnas.0803646105>
49. Nikolaou, N., Gordon, P. M., Hamid, F., Taylor, R., Lloyd-Jones, J., Makeyev, E. V., & Houart, C. (2022). Cytoplasmic pool of U1 spliceosome protein SNRNP70 shapes the axonal transcriptome and regulates motor connectivity. *Current Biology*, 32(23), 5099–5115e8. <https://doi.org/10.1016/j.cub.2022.10.048>
50. Madhu, R., Rodriguez, D., Guzik, C., Singh, S., De Tomaso, A. W., Valentine, M. T., & Loerke, D. (2020). Characterizing the cellular architecture of dynamically remodeling vascular tissue using 3-D image analysis and virtual reconstruction. *Molecular Biology of the Cell*, 31(16), 1714–1725. <https://doi.org/10.1091/mbc.E20-02-0091>
51. Ricci, L., Salmon, B., Olivier, C., Andreoni-Pham, R., Chaurasia, A., Alié, A., & Tiozzo, S. (2022). The onset of whole-body regeneration in *Botryllus schlosseri*: Morphological and molecular characterization. *Frontiers in Cell and Developmental Biology*, 10, 843775. <https://doi.org/10.3389/fcell.2022.843775>
52. Rosental, B., Raveh, T., Voskoboinik, A., & Weissman, I. L. (2020). Evolutionary perspective on the hematopoietic system through a colonial chordate: Allogeneic immunity and hematopoiesis. *Current Opinion in Immunology*, 62, 91–98. <https://doi.org/10.1016/j.coi.2019.12.006>
53. Yuan, L., Kurek, I., English, J., & Keenan, R. (2005). Laboratory-directed protein evolution. *Microbiology and Molecular Biology Reviews*, 69(3), 373–392. <https://doi.org/10.1128/MMBR.69.3.373-392.2005>
54. Rabinowitz, C., Douek, J., Weisz, R., Shabtay, A., & Rinkevich, B. (2006). Isolation and characterization of four novel thraustochytrid strains from a colonial tunicate. *Indian Journal of Marine Sciences*, 35(4), 341–350.
55. Shene, C., Garcés, M., Vergara, D., Peña, J., Claverol, S., Rubilar, M., & Leyton, A. (2019). Production of lipids and proteome variation in a Chilean *Thraustochytrium striatum* strain cultured under different growth conditions. *Marine Biotechnology*, 21(1), 99–110. <https://doi.org/10.1007/s10126-018-9863-z>
56. Qarri, A., Rinkevich, Y., & Rinkevich, B. (2021). Employing marine invertebrate cell culture media for isolation and cultivation of thraustochytrids. *Botanica Marina*, 64(6), 447–454. <https://doi.org/10.1515/bot-2021-0035>
57. Lyu, L., Wang, Q., & Wang, G. (2020). Cultivation and diversity analysis of novel marine thraustochytrids. *Marine Life Science & Technology*, 3(2), 263–275. <https://doi.org/10.1007/s42995-020-00069-5>
58. Sun, X.-M., Xu, Y.-S., & Huang, H. (2021). Thraustochytrid cell factories for producing lipid compounds. *Trends in Biotechnology*, 39(7), 648–650. <https://doi.org/10.1016/j.tibtech.2020.10.008>
59. Li, Y., Kim, J. I., Pysh, L., & Chapple, C. (2015). Four isoforms of *Arabidopsis* 4-coumarate:CoA ligase have overlapping yet distinct roles in phenylpropanoid metabolism. *Plant Physiology*, 169(4), 2409–2421. <https://doi.org/10.1104/pp.15.00838>
60. Hamberger, B., & Hahlbrock, K. (2004). The 4-coumarate:CoA ligase gene family in *Arabidopsis thaliana* comprises one rare, sinapate-activating and three commonly occurring isoenzymes. *Proceedings National Academy of Science USA*, 101(7), 2209–2214. <https://doi.org/10.1073/pnas.0307307101>
61. Hyams, Y., Rubin-Blum, M., Rosner, A., Brodsky, L., Rinkevich, Y., & Rinkevich, B. (2023). Physiological changes during torpor favor association with *Endozoicomonas* endosymbionts in the urochordate *Botryllodes leachi*. *Frontiers in Microbiology*, 14. <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1072053>
62. Ide, K., Nishikawa, Y., Maruyama, T., Tsukada, Y., Kogawa, M., Takeda, H., Ito, H., Wagatsuma, R., Miyaoka, R., Nakano, Y., Kinjo, K., Ito, M., Hosokawa, M., Yura, K., Suda, S., & Takeyama, H. (2022). Targeted single-cell genomics reveals novel host adaptation strategies of the symbiotic bacteria *Endozoicomonas* in *Acropora tenuis* coral. *Microbiome*, 10(1), 220. <https://doi.org/10.1186/s40168-022-01395-9>
63. Pogoreutz, C., & Ziegler, M. (2024). Frenemies on the reef? Resolving the coral-*Endozoicomonas* association. *Trends in Microbiology*, 11, 6. <https://doi.org/10.1016/j.tim.2023.11.006>
64. Schreiber, L., Kjeldsen, K. U., Funch, P., Jensen, J., Obst, M., López-Legentil, S., & Schramm, A. (2016). *Endozoicomonas* are specific, facultative symbionts of sea squirts. *Frontiers in Microbiology*, 7, 1042. <https://doi.org/10.3389/fmicb.2016.01042>
65. Alex, A., & Antunes, A. (2019). Comparative genomics reveals metabolic specificity of *Endozoicomonas* isolated from a marine sponge and the genomic repertoire for host-bacteria symbioses. *Microorganisms*, 7(12), Article 12. <https://doi.org/10.3390/microorganisms7120635>
66. Neave, M. J., Michell, C. T., Apprill, A., & Voolstra, C. R. (2017). *Endozoicomonas* genomes reveal functional adaptation and plasticity in bacterial strains symbiotically associated with diverse marine hosts. *Scientific Reports*, 7(1), Article 1. <https://doi.org/10.1038/srep40579>
67. Rees, M. A., Stinear, T. P., Goode, R. J. A., Coppel, R. L., Smith, A. I., & Kleefeld, O. (2015). Changes in protein abundance are observed in bacterial isolates from a natural host. *Frontiers in Cellular and Infection Microbiology*, 5, 71. <https://doi.org/10.3389/fcimb.2015.00071>
68. Knief, C., Delmotte, N., & Vorholt, J. A. (2011). Bacterial adaptation to life in association with plants: A proteomic perspective from culture to in situ conditions. *Proteomics*, 11(15), 3086–3105. <https://doi.org/10.1002/pmic.201000818>
69. Pérez, V., Hengst, M., Kurte, L., Dorador, C., Jeffrey, W. H., Wattiez, R., Molina, V., & Matallana-Surget, S. (2017). Bacterial survival under extreme UV radiation: A comparative proteomics study of *Rhodobacter* sp., isolated from high altitude wetlands in Chile. *Frontiers in Microbiology*, 8, 1173. <https://doi.org/10.3389/fmicb.2017.01173>
70. Worth, C. L., Gong, S., & Blundell, T. L. (2009). Structural and functional constraints in the evolution of protein families. *Nature Reviews Molecular Cell Biology*, 10(10), Article 10. <https://doi.org/10.1038/nrm2762>
71. Pearce, T. (2011). Evolution and constraints on variation: Variant specification and range of assessment. *Philosophy of Science*, 78(5), 739–751. <https://doi.org/10.1086/664568>
72. Wu, L., Candille, S. I., Choi, Y., Xie, D., Li-Pook-Than, J., Tang, H., & Snyder, M. (2013). Variation and genetic control of protein abundance in humans. *Nature*, 499(7456), 79–82. <https://doi.org/10.1038/nature12223>
73. Keele, G. R., Zhang, T., Pham, D. T., Vincent, M., Bell, T. A., Hock, P., Shaw, G. D., Paulo, J. A., Munger, S. C., Pardo-Manuel de Villena, F., Ferris, M. T., Gygi, S. P., & Churchill, G. A. (2021). Regulation of protein abundance

- in genetically diverse mouse populations. *Cell Genomics*, 1(1), 100003. <https://doi.org/10.1016/j.xgen.2021.100003>
74. Shorter, J., Couch, C., Huang, W., Carbone, M. A., Peiffer, J., Anholt, R. R. H., & Mackay, T. F. C. (2015). Genetic architecture of natural variation in *Drosophila melanogaster* aggressive behavior. *Proceedings National Academy of Science USA*, 112(27), E3555–E3563. <https://doi.org/10.1073/pnas.1510104112>
75. He, F., Steige, K. A., Kovacova, V., Göbel, U., Bouzid, M., Keightley, P. D., Beyer, A., & de Meaux, J. (2021). Cis-regulatory evolution spotlights species differences in the adaptive potential of gene expression plasticity. *Nature Communications*, 12(1), Article 1. <https://doi.org/10.1038/s41467-021-23558-2>
76. Zhan, T., Rindtorff, N., & Boutros, M. (2017). Wnt signaling in cancer. *Oncogene*, 36(11), 1461–1473. <https://doi.org/10.1038/ncr.2016.304>
77. Chen, W., & Moore, M. J. (2014). The spliceosome: Disorder and dynamics defined. *Current Opinion in Structural Biology*, 24, 141–149. <https://doi.org/10.1016/j.sbi.2014.01.009>
78. Deibel, D., McKenzie, C., Rise, M., Thompson, R. J., Lowen, J. B., Ma, K., Applin, G., O'Donnell, R., Wells, T., Hall, J., Sargent, P., & Pilgrim, B. (2014). Recommendations for eradication and control of non-indigenous, colonial, ascidian tunicates in newfoundland harbours. <https://www.semanticscholar.org/paper/Recommendations-for-Eradication-and-Control-of-%2C-%2C-Deibel-McKenzie/ba1857339dbe80fe8520c4323e9f7b5659bc3820>
79. Chase, A. L., Dijkstra, J. A., & Harris, L. G. (2016). The influence of substrate material on ascidian larval settlement. *Marine Pollution Bulletin*, 106(1–2), 35–42. <https://doi.org/10.1016/j.marpolbul.2016.03.049>
80. Wei, J., Wang, G., Li, X., Ren, P., Yu, H., & Dong, B. (2017). Architectural delineation and molecular identification of extracellular matrix in ascidian embryos and larvae. *Biology Open*, 6(9), 1383–1390. <https://doi.org/10.1242/bio.026336>
81. Lauzon, R. J., Ishizuka, K. J., & Weissman, I. L. (2002). Cyclical generation and degeneration of organs in a colonial urochordate involves crosstalk between old and new: A model for development and regeneration. *Developmental Biology*, 249(2), 333–348. <https://doi.org/10.1006/dbio.2002.0772>
82. Vizzini, A., Pergolizzi, M., Vazzana, M., Salerno, G., Di Sano, C., Macaluso, P., Arizza, V., Parrinello, D., Cammarata, M., & Parrinello, N. (2008). FACIT collagen (1alpha-chain) is expressed by hemocytes and epidermis during the inflammatory response of the ascidian *Ciona intestinalis*. *Developmental and Comparative Immunology*, 32(6), 682–692. <https://doi.org/10.1016/j.dci.2007.10.006>
83. Rinkevich, B., Tartakover, S., & Gershon, H. (1998). Contribution of morula cells to allogeneic responses in the colonial urochordate *Botryllus schlosseri*. *Marine Biology*, 131(2), 227–236. <https://doi.org/10.1007/s002270050315>
84. Rinkevich, B., & Rabinovitz, C. (2000). Urochordate cell cultures: From in vivo to in vitro approaches. In *Aquatic invertebrate cell cultures* (pp. 225–244). Springer-Praxis. <https://cir.nii.ac.jp/crid/1574231875003035264>
85. Rinkevich, B. (2011). Cell cultures from marine invertebrates: New insights for capturing endless stemness. *Marine Biotechnology (New York, N.Y.)*, 13(3), 345–354. <https://doi.org/10.1007/s10126-010-9354-3>
86. Hesp, K., van der Heijden, J. M. E., Munroe, S., Sipkema, D., Martens, D. E., Wijffels, R. H., & Pomponi, S. A. (2023). First continuous marine sponge cell line established. *Scientific Reports*, 13(1), Article 1. <https://doi.org/10.1038/s41598-023-32394-x>
87. Prünster, M. M., Ricci, L., Brown, F. D., & Tiozzo, S. (2019). Modular co-option of cardiopharyngeal genes during non-embryonic myogenesis. *EvoDevo*, 10(1), 3. <https://doi.org/10.1186/s13227-019-0116-7>

SUPPORTING INFORMATION

Additional supporting information may be found online <https://doi.org/10.1002/pmic.202300628> in the Supporting Information section at the end of the article.

How to cite this article: Kültz, D., Gardell, A. M., DeTomaso, A., Stoney, G., Rinkevich, B., Rinkevich, Y., Qarri, A., Dong, W., Luu, B., & Lin, M. (2024). Deep quantitative proteomics of North American Pacific coast star tunicate (*Botryllus schlosseri*). *Proteomics*, 24, e2300628. <https://doi.org/10.1002/pmic.202300628>