

# **Development of a gill assay library for ecological proteomics of threespine sticklebacks (*Gasterosteus aculeatus*)**

**Johnathon Li<sup>1</sup>, Bryn Levitan<sup>1</sup>, Silvia Gomez-Jimenez<sup>2</sup>, Dietmar Kultz<sup>1,3</sup>**

<sup>1</sup> Department of Animal Sciences, University of California Davis, Meyer Hall, One Shields Ave., Davis, CA 95616, USA

Email: [dkultz@ucdavis.edu](mailto:dkultz@ucdavis.edu), [joli@ucdavis.edu](mailto:joli@ucdavis.edu), [bblevitan@ucdavis.edu](mailto:bblevitan@ucdavis.edu)

Phone: (530) 752-2991

<sup>2</sup> Centro de Investigación en Alimentación y Desarrollo, Carretera a la Victoria Km. 0.6, Apdo. Postal # 1735., Hermosillo, Sonora, México C.P. 83000

Email: [s.gomez@ciad.mx](mailto:s.gomez@ciad.mx)

Phone: (662) 289-24-00 Ext. 334

<sup>3</sup> Coastal Marine Sciences Institute, University of California, Davis,

**Correspondence:** Dietmar Kultz, Professor, ANS Department, University of California Davis, 2251 Meyer Hall, One Shields Ave., Davis, CA 95616, USA; [dkultz@ucdavis.edu](mailto:dkultz@ucdavis.edu); (530) 752-299

**Running title:** Gill assay library for ecological proteomics

## *Abbreviations*

ACN	acetonitrile
BodHar	Bodega Harbor
CAMP	comparative animal molecular phenotyping
CARD	caspase recruitment domain
csv	comma separated value
CLCN2	chloride channel protein 2
DDA	data-dependent acquisition,
DIA	data-independent acquisition
dotp	dot product value
FA	formic acid
FDR	false discovery rate
FC	fold change
ECM	extracellular matrix
HRM	hyper-reaction monitoring
iRT	internal retention time
LaBoRo	Laguna de la Bocana del Rosario
LakSol	Lake Solano
MRM	multiple-reaction monitoring
Mx	interferon-induced GTP-binding protein Mx
PS	population sampling
ppm	parts per million
PRM	parallel reaction monitoring
PTMs	post translational modifications
PYCARD	apoptosis associated speck-like protein containing a CARD
SWATH	sequential windowed acquisition of all theoretical fragment ion mass spectra
TIC	total ion current
TPP	transproteomics pipeline
UHR	ultra high resolution
UPLC	ultra performance liquid chromatography
WesLag	Westchester Lagoon

## Summary

A data-independent acquisition (DIA) assay library for quantitative analyses of proteome dynamics has been developed for gills of threespine sticklebacks (*Gasterosteus aculeatus*). A raw spectral library was generated by data-dependent acquisition (DDA) and annotation of tryptic peptides to MSMS spectra and protein database identifiers. The assay library was constructed from the raw spectral library by removal of low-quality, ambiguous, and low-signal peptides. Only unique proteins represented by at least two peptides are included in the assay library, which consists of 1,506 proteins, 5,074 peptides, 5,104 precursors, and 25,322 transitions. This assay library was used with DIA data to identify biochemical differences in gill proteomes of four populations representing different eco- and morpho-types of threespine sticklebacks. The assay library revealed unique and reproducible proteome signatures. Warm-adapted, low-plated, brackish-water fish from Laguna de la Bocana del Rosario (Mexico) show elevated HSP47, extracellular matrix, and innate immunity proteins whereas several immunoglobulins, interferon-induced proteins, ubiquitins, proteolytic enzymes, and nucleic acid remodeling proteins are reduced. Fully-plated, brackish-water fish from Westchester Lagoon (Alaska) display elevated ion regulation, GTPase signaling, and contractile cytoskeleton proteins, altered abundances of many ribosomal, calcium signaling and immunity proteins, and depleted transcriptional regulators and metabolic enzymes. Low-plated freshwater fish from Lake Solano (California) have elevated inflammasomes and proteolytic proteins while several iron containing and ion regulatory proteins are reduced. Gills of fully-plated, marine fish from Bodega Harbor (California) have elevated oxidative metabolism enzymes and reduced transglutaminase 2, collagens, and clathrin heavy chains. These distinct proteome signatures represent targets for testing ecological and evolutionary influences on molecular mechanisms of gill function in threespine sticklebacks. Furthermore, the gill assay library represents a model for other tissues and paves the way for accurate and reproducible network analyses of environmental context-dependent proteome dynamics in complex organisms.

## Introduction

Enormous efforts have been devoted towards development of quantitative proteomics approaches that enable large network-scale analyses of proteome dynamics. As a result, it is now possible to routinely assess systems-scale protein dynamics with superior quantitative accuracy, e.g. for quality control of antibody selectivity and specificity (1). An exciting recent development is data-independent acquisition (DIA) technology, which permits precise and simultaneous quantitation of thousands of proteins (2, 3). This technique is well suited for label-free approaches to be used with virtually any species studied in its natural environment (4, 5). However, application of these technical developments for ecological proteomics is not common and its potential is underutilized. One of the main technical challenges and time-consuming aspects for precise DIA quantitative proteomics is the selection and validation of diagnostically reliable transitions that unambiguously and quantitatively represent unique proteins in diverse samples (6-8). In this study we have addressed this challenge by developing a reliable, standardized, and manually curated DIA assay library for gills of threespine sticklebacks (*Gasterosteus aculeatus*). DIA targeted data extraction has recently been introduced for biomedical applications and shown to be as quantitatively consistent and accurate as selected reaction monitoring (SRM), which is considered the gold standard of quantitative proteomics (3, 7, 8). DIA-based assays combine advantages of selective/multiple reaction monitoring (SRM/ MRM) approaches on the one hand and large-scale, high-throughput DDA on the other hand. Like SRM/ MRM approaches, DIA methods monitor well-defined transitions representing specific precursors and proteins, although, in DIA the transitions are targeted post-acquisition by utilizing MSMS spectral libraries (9). In addition, DIA approaches are suitable for quantitation of more than a thousand proteins simultaneously, which is why they are sometimes referred to as hyper-reaction monitoring (HRM) (10). The utility of the stickleback gill assay library developed in this study was demonstrated by comparing gill proteomes of four stickleback populations adapted to four different environments. Quantitative reproducibility of this assay library was verified by analyzing four different population samplings for each of the four habitats.

Threespine sticklebacks are euryhaline fish that represent a long-standing model in behavioral, evolutionary, and ecological physiology research (11). These fish are ancestrally marine but the species is also represented by many anadromous and land-locked populations that have colonized brackish and freshwater habitats following the retreat of Pleistocene glaciers (12-14). This species is eurythermal but typically restricted to cold-temperate climates throughout coastal areas in the entire Northern hemisphere (15, 16). However, some exceptional refuge populations have evolved to colonize warm-water habitats (17). These refuge populations provide a unique opportunity for analyzing the genetic and biochemical (i.e. proteomic) mechanisms underlying warm-adaptation in vertebrates. Such studies inform us about biochemical coping strategies of vertebrates responding to a changing and stressful climate. Because vertebrates have long generation times, experimental evolution approaches (18) are not feasible. Instead, such studies rely on naturally evolved populations to reveal the mechanisms of environmental adaptation on an evolutionary time scale.

Threespine sticklebacks ideally combine key advantages as a model organism for integrative biology and cross-disciplinary research at the interface of ecophysiology, behavior, evolutionary biology, biochemistry and genetics. They are represented by many populations that have naturally adapted to diverse environments, their complete genome has been sequenced, and their proteome is well annotated (14). A sequenced and well-annotated genome represents a major resource for studying the biology of a particular species (19). Therefore, much attention has been devoted towards developing and applying whole genome sequencing approaches and strategies for genome, transcriptome, and proteome annotation to an increasing number of species (20). The ability to unambiguously map peptide and protein sequences based on a well-annotated species-specific reference database to unique genomic loci represents a critical prerequisite for proper proteome-wide and unambiguous quantitative analyses. In this study we have utilized these unique resources and advantages of threespine sticklebacks to develop a strategy for long-term quantitative ecological proteomics of this species. This strategy is based on the generation of validated assay libraries that are tissue-specific and can be used in a highly consistent manner, i.e. allowing the quantitation of the same set of proteins based on identical, previously validated precursors

and transitions in every sample. To demonstrate its usefulness, the current study has been limited to a single tissue (gill), but the approach can be expanded to include other tissues for integrative studies of environmental effects on whole organism proteomes. The availability of this gill assay library will propel precise, quantitative and network-scale analyses of proteome dynamics in fish experiencing a wide variety of ecological and evolutionary contexts.

## **Experimental Procedures**

### ***Experimental Design and Statistical Rationale***

Population samplings (PSs) were performed at Laguna de la Bocana del Rosario, Mexico, Lake Solano, California, Bodega Harbor, California, and Westchester Lagoon, Alaska. Each PS is represented by six biological replicates (individual fish) from each of the four habitats (24 fish total per PS, see below for population habitat characterization). This experimental design was repeated four times to enable replicated and cumulative population comparisons (PS1, PS2, PS3, PS4 yielding 24 fish per population and 96 fish total). For each fish, gill proteins were extracted and analyzed using procedures outlined below such that each sample analyzed represents an independent biological replicate. For PS1 all 24 samples were processed both by DDA (to create the spectral and assay libraries) and DIA. The remaining 72 samples collected in PS2, PS3, and PS4 were only processed by DIA to assess the robustness of gill proteome population differences detected. Statistical analyses of quantitative DIA results was performed with MSstats, which uses mixed linear models that have been optimized for the quantitative analysis of label-free DIA data in Skyline (21, 22). A power analysis was performed using MSstats and DIA data obtained from PS1 to obtain confidence intervals and determine fold difference thresholds of 2.0 for separate PS comparisons ( $n = 6$  per population) and 1.45 for the cumulative comparison ( $n = 24$  per population). The false discovery rate (FDR) was set to 1% for protein identification from DDA data. For peak scoring of DIA data, mProphet (23) was used in combination with an equal number of decoy peptides as assay peptides and an FDR of  $<0.5\%$ . For MSstats (22) statistical analyses of DIA data the

confidence interval threshold was set to >99% and the Benjamini-Hochberg (24) multiple testing-corrected significance threshold was  $p < 0.05$ .

### ***Sample collection and preparation***

Threespine sticklebacks (*Gasterosteus aculeatus*) were collected from four different habitats using seining and traps as previously described (4). The four habitats were purposely chosen to include populations that represent diverse ecotypes and morphotypes and are adapted to different temperatures and salinities. Thus, the spectral library generated in PS1 is representative of low-plated, warm-adapted, brackish water fish from Laguna de la Bocana del Rosario (Baja California, Mexico), fully plated, brackish/anadromous fish from Westchester Lagoon (Anchorage, Alaska), low-plated, land-locked freshwater fish from Lake Solano (Winters, California), and fully plated, resident marine fish from Bodega Harbor (Bodega Bay, California). For brevity these four populations are referred to as LaBoRo, WesLag, LakSol, and BodHar respectively. Their sampling locations and habitat conditions, morphotype, and ecotype are summarized in figure S1. Additional phenotypes for each fish are provided in figure S2. To equally represent both sexes in the spectral library and assay library, three males and three females were included for each population in PS1. Gill dissection, protein extraction, and protein digestion with immobilized trypsin were performed using an optimized sample preparation protocol as described previously (25).

### ***Data-dependent acquisition (DDA)***

Tryptic peptides (2  $\mu\text{L}$ , 300 ng/  $\mu\text{L}$ ) were separated along a 5 to 30% acetonitrile (ACN) gradient over 120 min using a nanoAcquity (Waters, Milford, MA) UPLC and injected into an ImpactHD UHR qTOF mass spectrometer (Bruker Daltonics, Bremen, Germany). The conditions used for DDA were as follows: Peptides were first trapped for one min (10  $\mu\text{L}/\text{min}$  flow rate) on a trap column (Symmetry, Waters 186003514) and then separated on a 1.7  $\mu\text{m}$  particle size BEH C18 column (250 mm, 75  $\mu\text{m}$ , Waters 186003545) by reversed phase chromatography. The aqueous solvent contained 0.1% formic acid (FA), which was omitted from the organic phase to prevent formation of brown ACN aggregates that would

otherwise precipitate at low pH at the pico-emitter tip (New Objective FS360–20-10-D-20). Low concentration tuning mix (Agilent G1969–85000) was used daily for mass calibration of the ImpactHD instrument. Batch-processing of samples was controlled with Hystar 4.1 software (Bruker Daltonics). Peak lists were generated with DataAnalysis 4.4 (Bruker Daltonics), imported into PEAKS 8.0 (Bioinformatics Solutions Inc., Waterloo, Canada) and peptide to spectrum matches identified using three search engines (PEAKS 8.0, Mascot 2.2.7, and XTandem Cyclone) followed by unambiguous assignment of peptides to unique proteins and the *G. aculeatus* proteome database downloaded from UniprotKB, version June 27, 2015 (5). The database searched included 27,248 *G. aculeatus* proteins plus the same number of randomly scrambled decoys and 282 common contaminants (human keratins, porcine trypsin, etc.). Trypsin cleavage specificity was C-terminal of either Lys or Arg except when followed by Pro and a maximum of two missed cleavages were allowed. First round search variable PTMs allowed were Cys carbamido-methylation, Met oxidation, Protein N-terminal acetylation and Pro hydroxylation. Second round PEAKS-PTM searches included all 313 variable PTMs contained in the PEAKS database (max. 3 PTMs per peptide). Mass tolerance limits were 0.02 Da for precursors and 20 ppm for fragment ions. PEAKSQ Top3 MS1 quantitative profiling (table S1) and Scaffold 4.8 (Proteome Systems, Portland, OR) spectral counting (table S2) were performed on PS1 DDA data as previously described (5). Complete data, metadata, and results of the MS1 level surveys performed for PS1 are available at CAMP proteome (AC CAMPDDA00023, [https://kueltzlab.ucdavis.edu/CAMP\\_dda\\_profiles.cfm?AC=CAMPDDA00023](https://kueltzlab.ucdavis.edu/CAMP_dda_profiles.cfm?AC=CAMPDDA00023)). Data and metadata for these DDA analyses, including the Scaffold file, have been submitted to MassIVE (MSV000081795) and ProteomeXchange (PXD008395) and are publicly accessible under MassIVE project title: Identification and spectral annotation of the threespine stickleback gill proteome.

### ***Data-independent acquisition (DIA)***

DIA was performed on all samples collected in all four PSs (96 total biological replicates) using the same instrumentation and separation conditions as for DDA, nanoAcquity (Waters) interfaced online to an ImpactHD UHR qTOF (Bruker Daltonics). Slight variations in retention time (RT) were compensated for



by application of internal retention time (iRT) standard calibration in Skyline (figure one). DIA conditions reported previously were used (26) except that the isolation width was set to 10 m/z (0.5 m/z overlap) and spectra were collected over 75 isolation width windows ranging from 390 to 1065 m/z at 37.5 Hz and in two sec. intervals. Except for the acquisition frequency and operating the mass spectrometer in MRM mode, all acquisition parameters were identical to those used for DDA. DIA data from PS1 were used to filter the target list and for selecting iRT standards with Skyline 3.6 (21). Overall system performance and quantitative reproducibility over time was verified weekly by DIA of an aliquot of the same sample (a gill sample from a LakSol stickleback). If necessary, system performance was restored by cleaning the capillary, hexapole cartridge, quadrupole, and collision cell of the mass spectrometer. DIA data and metadata for all four PSs, including all transition, peptide, and document parameters, the final target list, and the *G. aculeatus* gill spectral library have been deposited and are freely accessible at Panorama Public (<https://panoramaweb.org/labkey/JoLi-01.url>).

### ***Spectral library and DIA assay library construction***

A *G. aculeatus* gill spectral library was generated using the peptide-to-spectrum matches and protein annotation information generated from PS1 DDA data. This information was exported in pepxml and Mascot dat formats and then imported into Skyline 3.6 (21). Skyline was used for spectral library construction and for generating an initial target list, which contained 400,412 transitions, 60,132 precursors, 50,501 peptides, and 5,998 proteins (figure two a-d). Transitions for the initial target list were automatically chosen from library spectra using the Skyline transition settings dialogue based on the following criteria: ion 3 to last ion -1; exclusion of the precursor isolation window (m/z width = 10); charge 1, 2; precursor charge 2-4; mass accuracy threshold within  $\pm 0.035$  m/z of the expected mass. This initial target list was filtered sequentially in seven steps using Skyline. The impact of each filtering step on the number of transitions, precursors, peptides, and proteins lost from the assay library is shown in figure two a-d. Filter step one consisted of exclusion of PTMs other than Met oxidation and Cys carbamido-methylation. During filter step two all repeated peptides were excluded and peptide

uniqueness by proteins was enforced. Filter step three excluded all precursors with less than four transitions and filter step four excluded all peptides with any missed cleavages. The remaining three filter steps were performed using a sample training set, which consisted of all 24 PS1 samples (six from each of the four populations). All training samples were subjected to iRT calibration and mProphet (23) peak detection using Skyline as follows: First, iRT calibration was performed based on 16 internal standard peptides (Figure one a, for iRT standard peptides see Panorama Public, <https://panoramaweb.org/labkey/JoLi-01.url>). The Skyline mProphet (23) module was then used for optimizing peak detection and scoring after generating randomly scrambled decoy peptides equal in number to assay peptides. The q value threshold for mProphet peak detection was set at 0.005 (FDR < 0.5%). After mProphet peak detection the iRT calibration was repeated to correctly reflect the RT of optimized peaks. After iRT calibration and mProphet peak scoring for the 24 PS1 samples it was possible to apply three additional filter steps that considered the quality of peaks for each peptide in the assay library. The advanced refine feature of Skyline was used in step five to exclude all peptides with dotp values <0.8 in all 24 samples and in step six to exclude all proteins that were represented by less than two remaining peptides. The final filter step (seven) consisted of curation of the assay library by manual inspection of the peak shapes for all peptides in the 24 PS1 samples. This step was performed independently by two investigators and resulted in removal of remaining low intensity peptides and large interferences. The rules followed for manual curation were as follows: 1. If an interference was greater than the actual peak at proper RT then the corresponding transition was removed; 2. If a precursor peak is not clearly discernable in at least half the samples of one population within three min of the predicted RT then the precursor was removed; 3. If precursor peak widths differ by more than two-fold for half of the samples of the same population from the other half, then the precursor was removed; 4. If more than one transition of the five nominal transitions shows an interference that exceeds the intensity of the actual transition peak at the predicted RT then the corresponding precursor was removed. The resulting final target list contains 25,322 transitions, 5,104 precursors, 5,074 peptides, and 1,506 proteins (figure one a-

d). The complete assay library including all relevant metadata is available at Panorama Public (<https://panoramaweb.org/labkey/JoLi-01.url>). This assay represents a tier two assay (27).

### ***Statistical analyses and visual data representation***

Quantitative DIA data were normalized at the MS2 level by equalizing medians in each sample followed by comparison of protein quantities in each population against all other populations with the MSstats (22) module integrated in Skyline using the following parameters: confidence interval > 99% (FDR < 1%), scope = protein, summary method = Tukey's median polish, use zero for missing peaks, and q value < 0.005. The q value represents the FDR associated with the mProphet peak scoring model applied to the DIA data, i.e. mProphet FDR < 0.5%. The manual curation was the final step (step seven) after mProphet peak scoring during library assay generation. This final step was only performed for PS1 samples, which were used as training samples for steps five to seven of assay library construction (see previous paragraph). Once the assay library had been established using PS1 samples no manual curation was performed for subsequent DIA data sets (PS2, PS3, and PS4). Rather, the same assay library, parameters, and thresholds as used for PS1 were applied in conjunction with iRT calibration and mProphet peak scoring to quantify transition peaks for all subsequent DIA data sets (PS2, PS3, and PS4). Reproducibility of proteins identified as significantly different in abundance was assessed using two common multiple testing-corrected p value thresholds (0.01 and 0.05). This analysis demonstrates that a multiple testing-corrected p value threshold of 0.05 represents the best compromise for balancing false positives with false negatives (figure three). DIA data exported from Skyline were used for generating heat maps with Genesis 1.8.1 (28). The Venny 2.1 tool was used for visual representation of data as Venn diagrams (29).

### ***Pathway and gene ontology analyses***

The UniprotKB ID mapping tool was used to annotate stickleback gill proteins with their corresponding KEGG (Kyoto Encyclopedia of Genes and Genomes <https://www.kegg.jp/>) orthology (KO) identifiers. KO identifiers were assigned to 1307 of the 1506 proteins contained in the final assay library (table S3). Sets of proteins that were significantly enriched/ depleted in each of the four populations were then

subjected to pathway analysis using the KEGG Mapper tool ([https://www.genome.jp/kegg/tool/map\\_pathway1.html](https://www.genome.jp/kegg/tool/map_pathway1.html)). This approach identified corresponding KEGG pathway maps that contain proteins for each of the population-specific enriched/ depleted sets. Pathway maps containing 5 proteins or more per set were considered most informative. PANTHER (Protein Annotation Through Evolutionary Relationships) was utilized for gene ontology (GO) and pathway enrichment analysis as previously described (4, 5, 30). A hidden Markov model (HMM) based gene ontology PANTHER library for the entire threespine stickleback proteome constructed on January 25, 2015 was used for this purpose (5). This library contains the PANTHER IDs for all but 12 of the 1506 gill assay library proteins (table S3). Stickleback PANTHER IDs of population-enriched/depleted protein sets were searched against the complete assay library of 1506 proteins to identify significantly enriched and depleted GO categories and pathways with the PANTHER enrichment test (Release date: April 13, 2018). A multiple-testing corrected PANTHER enrichment test  $p < 0.05$  was considered statistically significant for enrichment analyses.

## Results

### *Gill proteome identification and spectra annotation*

Unbiased DDA proteome profiling of all samples collected in PS1 followed by PEAKS 8.0 (Bioinformatics solutions), Mascot 2.2.7 (Matrix Science Inc.), and X!Tandem Cyclone (the GPM) database searching unambiguously identified 2,101 unique proteins that are represented by 17,568 unmodified peptides (MassIVE AC MSV000081795). In addition, label-free quantitation of protein abundances using the MS1-Top3 approach with PEAKSQ (determination of chromatographic peak area for the three most abundant, unique MS1 peptides) yielded many candidate proteins that are differently abundant in the four populations (table S1, CAMP proteome AC CAMPDDA00023). After consolidating database search results from multiple search engines (PEAKS, Mascot, and X!Tandem) in Scaffold 4.8 (Proteome Systems Inc.) the complete dataset contained 1652 proteins represented by at least two unique peptides at 1% protein FDR and 0.1% peptide FDR (table S2, CAMP proteome AC CAMPDDA00023).

The Scaffold file also contains semi-quantitative spectral count analyses that suggest candidate proteins whose abundances are population-specific (MassIVE AC MSV000081795, CAMP proteome AC CAMPDDA00023). The PEAKS Top3 and Scaffold spectral counting results obtained from PS1 indicate two broad trends: First, gill proteomes differ markedly between the four populations. Second, the gill proteomes of the LaBoRo and WesLag populations are most unique compared to the other populations. For the purpose of this study the DDA data were used for assay library generation and not for systematic comparison of top3 and spectral counting approaches with assay library quantitation, which exceeds the scope of this article.

### ***Generation of the spectral library and cDIA assay library***

Search results for DDA data of all 24 PS1 samples were obtained by multiple search engines, consolidated with PEAKS Studio, and used for MSMS spectral library construction. A highly diverse set of samples was purposely collected from different populations, morphotypes (fully and low plated), and ecotypes (resident marine, resident freshwater, and brackish water) of *G. aculeatus* adapted to different temperature, salinity, and latitude (figure S1). This sample diversity supports a wide representation of pertinent *G. aculeatus* proteins in the spectral library and empowers systematic future analyses of differentially abundant proteins in diverse populations.

The initial spectral library contains a much greater number of transitions, peptides, precursors, and proteins than what is included in the final assay library (figure two a-d). This large difference is due to redundancy, ambiguity, and low-quality targets that are removed during target list filtering. DIA was performed on all 24 PS1 samples to assess the two broad trends in population-dependent proteome dynamics outlined above more precisely. The resulting PS1 DIA dataset was used as a training sample set for target list filtering to generate the final assay library. The numbers of peptides and precursors representing each protein included in the final, filtered assay library are depicted in figure two e. Most proteins are represented by three precursors/ peptides and the average number of precursors/ peptides per protein in this assay library is 3.9. Almost all precursors in the assay library (4881) are represented by the

maximum number of five transitions with the remainder being represented by four transitions. Almost all peptides in the assay library are represented by a single charge state and only 15 peptides are represented by two precursors. The most common precursor charge in the assay library is 2+ (4118 precursors) followed by 3+ (870 precursors), 4+ (66 precursors), and 1+ (47 precursors). The most frequent transition ions in the assay library are y4+, y5+, y6+, y7+, y8+, and b3+ (figure two f). The complete optimized *G. aculeatus* gill assay library can be downloaded from Panorama Public (<https://panoramaweb.org/labkey/JoLi-01.url>) and CAMP proteome ([https://kueltzlab.ucdavis.edu/CAMP\\_dia\\_exp.cfm?ac=CAMPDIA00001](https://kueltzlab.ucdavis.edu/CAMP_dia_exp.cfm?ac=CAMPDIA00001)).

### ***Reproducibility of population-specific proteome signatures***

Power analysis performed with MSstats yielded a fold change (FC) threshold of 2.0 when each PS was analyzed separately (n = six per population) and 1.45 when all four PSs were combined (n = 24 per population, figure three a). The number of false positives and false negatives was estimated at the FC threshold of 2.0 at two different multiple-testing corrected alpha levels ( $p < 0.05$  and  $p < 0.01$ ). Proteins that exceed this FC of 2.0 in any of the four separate PSs (PS1, PS2, PS3, PS4) were graphed on volcano plots for each population and each PS (figure three b). The false positive rate was estimated as the number of proteins from this subset that display the opposite direction of regulation in any of the four PS at significance thresholds of  $p < 0.05$  and  $p < 0.01$ . The false negative rate was estimated as the number of proteins from this subset that exceed the FC threshold of 2.0 in the correct direction but do not reach the significance threshold ( $p < 0.05$  and  $p < 0.01$ ) in a particular PS. The number of false positive proteins defined in this way is very low. More importantly, the increase in the number of such false positive proteins is very small when the significance threshold is changed from  $p < 0.01$  to  $p < 0.05$  (figure three c). In contrast, the number of false negative proteins is much higher, and it increases greatly when the significance threshold is changed from  $p < 0.05$  to  $p < 0.01$  (figure three d). Therefore, and because correction for multiple testing already treats p values conservatively, a significance threshold of  $p < 0.05$  was applied to quantitative population comparisons performed with the DIA assay library.

The four repeated PSs enabled validation of proteins with significantly and at least two-fold different abundances in a single population relative to the other populations. Data from all four PSs had a mass error less than 10 ppm for a great majority and less than 20 ppm for all MS2 spectra collected from all samples (figure one). Retention time reproducibility was also very high for all samples in all four PSs as illustrated by the elution profiles of iRT standards (figure one). General trends observed in PS1 and the magnitude of differences between populations as illustrated by volcano plots were reproducible in the other three PSs (figure three). The only exception is the BodHar population for which much fewer abundance differences have been identified in PS1 and PS4 than in PS2 and PS3 (figure three). Volcano plots illustrate that some differences between PS regarding the sets of proteins identified as significantly differentially abundant in each of the populations are to be expected. They show that many of these proteins cluster near the two-fold threshold, which is exceeded in some but not all PSs (figure three). Almost all proteins that were significantly altered in a population in one PS were also altered in the same FC direction in the other PSs (figure three b, table S4). Even though statistical significance ( $n = \text{six per population}$ ,  $p < 0.05$ ,  $FC \geq 2.0$ ) was not reached for many of these proteins in all four PS, highly reproducible proteome signatures are evident for all populations. The LaBoRo proteome signature consists of 33 proteins that are significantly elevated and 23 proteins that are significantly reduced in all four PSs (table S4). The LakSol proteome signature consists of six proteins that are significantly elevated and two proteins that are significantly reduced in all four PSs (table S4). The WesLag proteome signature consists of 30 proteins that are significantly elevated and 29 proteins that are significantly reduced in all four PSs (table S4). The BodHar proteome signature consists of one protein (glutamine synthetase) that is significantly elevated and one protein (transglutaminase 2) that is significantly reduced in all four PSs (table S4). In summary, even with a relatively low number of biological replicates (six fish per population) the four PSs illustrate that robust population-specific proteome signatures consisting of many reproducibly statistically significant proteins can be obtained with the DIA assay library approach. In addition, these proteome signatures illustrate that LaBoRo and WesLag gill proteomes differ most from those of the other populations while the BodHar proteome signature is the least unique.

### ***Biological relevance of population-specific proteome signatures***

Samples from all four PSs were pooled into a single comparison to identify more comprehensive proteome signatures that are specific for each of the four populations. Pooling samples reduces the FC threshold from 2.0 to 1.45 at the same statistical power (figure three a). Analysis of the combined set of samples including all four PSs with MSstats yields 258 proteins that are significantly elevated and 242 proteins that are significantly reduced in LaBoRo fish (figure four, table S5). The WesLag population is characterized by 293 significantly elevated and 406 significantly reduced proteins (figure four, table S5). The other two populations have fewer significantly elevated (231 for LakSol and 178 for BodHar) and reduced (145 for LakSol and 207 for BodHar) proteins (figure four, table S5). The corresponding volcano plots verify that there is no bias in overall protein abundance for any of the populations. They also show that gill proteomes of the LaBoRo and WesLag populations are more unique (i.e. proteins have larger FC and higher statistical significance) than those of LakSol and BodHar populations (figure four).

KEGG pathway and PANTHER GO term and pathway enrichment analyses revealed certain pathways and biological functions that are associated with population-specific proteome signatures. Many KEGG pathways and PANTHER GO terms and pathways are represented by the 1506 protein included in the assay library (tables S5 and S6). PANTHER enrichment analyses using the population-specific sets of significantly elevated and reduced proteins against the complete assay library set of 1506 proteins reveals that the integrin signaling pathway is depleted ( $p = 0.0301$ ) among elevated LaBoRo proteins while nucleic acid binding ( $p = 0.00734$ ), RNA binding ( $p = 0.00357$ ), DNA binding ( $p = 0.0349$ ), translation regulator activity ( $p = 0.00364$ ), and translation factors ( $p = 0.001$ ) are enriched among reduced LaBoRo proteins (figure five a, b). Moreover, KEGG pathways represented by at least five proteins elevated in the LaBoRo population indicate a remodeling of extracellular matrix, cell adhesion, plasma-membrane anchored cytoskeleton, and the complement system (figure five a). Different apoptosis proteins are represented in KEGG pathways associated with elevated and depleted LaBoRo protein sets suggesting that apoptosis is also remodeled in this population. Furthermore, proteasome, lysosome, ribosome, amino



acid synthesis, RNA processing, antigen processing and presentation, and necroptosis pathways are represented by sets of at least five proteins that are reduced in the LaBoRo population indicating that protein turnover and adaptive immunity are reduced (figure five b). Pathways that are uniquely regulated in LaBoRo fish are reflected in the most highly and significantly elevated and reduced proteins. Two HSP47 isoforms, many extracellular matrix proteins, complement proteins, and cell adhesion regulatory proteins are most highly and significantly elevated while several interferon-induced proteins, immunoglobulins, ubiquitins, proteases, DNA/RNA processing proteins, and peptidyl prolyl cis/trans isomerases are most reduced in LaBoRo fish (figure six).

PANTHER enrichment analyses for the set of elevated LakSol proteins shows that hydrolase ( $p = 0.00129$ ) and peptidase ( $p = 0.00257$ ) activities are enriched while nucleic acid ( $p = 0.00187$ ) and RNA ( $p = 0.00136$ ) binding proteins are depleted (figure five c). KEGG pathways represented by at least five proteins that are elevated in the LakSol population indicate proteostasis and immunity related processes (ribosome, lysosome, phagosome, antigen processing/ presentation, apoptosis, necroptosis) are enriched (figure five c). KEGG pathways represented by reduced LakSol proteins include focal adhesion, ECM-receptor interaction, tight junctions, PI3K-Akt signaling, actin cytoskeleton regulation, and oxidative metabolism (figure five d). The most highly and significantly elevated LakSol proteins are caspase 1/interleukin 1 beta convertase, apoptosis-associated speck-like protein containing a CARD (PYCARD), cathepsins, and other proteases (figure seven a), which is consistent with the most prominent pathways in this set. However, the most significantly and highly reduced LakSol proteins are indicative more of iron transport and ion regulation than the KEGG pathways that are numerically most highly represented in the set of reduced LakSol proteins (figure seven b).

PANTHER enrichment analysis did not yield any significant biological processes or pathways for the sets of elevated and reduced BodHar proteins. However, several KEGG pathways involved in oxidative and cholesterol metabolism and growth are well represented in the set of elevated BodHar proteins while cell adhesion, immune-related pathways, apoptosis, necroptosis, and ECM-related KEGG pathways are prominent in the set of reduced BodHar proteins (figure six e, f). The most significantly and highly

elevated BodHar proteins represent the above mentioned KEGG pathways of oxidative and lipid metabolism but at least two of these proteins are also indicative of increased demands for long-term osmotic homeostasis in SW (glutamine synthetase and antifreeze protein LS12, figure seven c). Likewise, several proteins that are most highly reduced in BodHar fish, including collagens and proteins involved in cell adhesion and pattern recognition such as transglutaminase 2, clathrins, and alpha-2-macroglobulin represent the corresponding KEGG pathways mentioned above (figure seven d). Overall, the extent of abundance differences is lower for elevated and reduced BodHar proteins than for other populations (figures four and seven).

PANTHER enrichment analyses for the set of elevated WesLag proteins shows that ribonucleoprotein complex ( $p = 0.0379$ ), hydrolase activity acting on ester bonds ( $p = 0.0385$ ) and phosphatase activity ( $p = 0.027$ ) are enriched while integrin signaling ( $p = 0.0341$ ) and transcription ( $p = 0.0404$ ) are depleted (figure five e). For the set of reduced WesLag proteins PANTHER analysis indicates enrichment of amino acid metabolism ( $p = 0.0364$ ) and blood coagulation ( $p = 0.00426$ ) but depletion of organelle organization ( $p = 0.0464$ ), transcription factors ( $p = 0.0471$ ) and defense/ immunity proteins ( $p = 0.0028$ ) (figure five f). KEGG pathways well represented in the set of elevated WesLag proteins include calcium and GTPase signaling pathways, regulation of actin cytoskeleton, focal adhesion, tight junction, ECM receptor interaction, and proteostasis/ protein turnover (figure five e). Oxidative metabolism, complement and coagulation cascades, RNA transport, endocytosis, adrenergic signaling and phagosome are represented in both elevated and reduced WesLag protein sets, suggesting that these processes are remodeled relative to the other populations. However, there are also KEGG pathways that are only prominent in the set of depleted WesLag proteins including endocannabinoid and PPAR signaling, spliceosome, lysosome, antigen processing and presentation, necroptosis, glutathione metabolism, and pentose phosphate pathway (figure five f). KEGG pathways that are prominent in the set of elevated WesLag proteins are exemplified by the most highly and significantly elevated WesLag proteins, including many calcium regulatory proteins and proteins involved in GTP signaling, immunity, cell adhesion, cytoskeletal organization, pattern recognition, and protein translation. In addition, these proteins include ion transporters such as a

chloride channel and two subunits of the Na<sup>+</sup>/K<sup>+</sup>-ATPase (figure eight a). The most highly reduced and significant WesLag proteins mostly reflect the prominent KEGG pathways in this set and include immunity proteins such as complement proteins, thymosin, and MHC proteins, proteins involved in oxidative metabolism, and transcriptional regulators (figure eight b).

## **Discussion**

### ***Assay library development***

DIA targeted data extraction was first introduced six years ago (3). This method eliminates common problems with protein quantitation by DDA (stochastic, irreproducible selection and undersampling of precursor ions) while greatly exceeding the scope (number of proteins) that can be quantified by targeted approaches (SRM, MRM, PRM). Since its inception, many variants of DIA technology have been employed. Two major variants include SWATH on triple-quad instruments and DIA on UHR-qTOF instruments (9, 31-35). Recently, DIA targeted data extraction has been advanced and quantitative assay precision improved by providing tools that permit target list filtering and generation of curated target lists based on stringent quality control criteria, some of which are evaluated using a “training sample set”. In this study we have used Skyline (21) for targeted DIA data extraction and filtering the target list. The resulting assay library is much shorter than the initial raw target list generated from spectral libraries. Quantitative precision of the assay library is greatly improved because protein assignment ambiguities, low-quality and low-signal transitions, and proteins with less than two unique precursors were eliminated during target list filtering. However, target list filtering also eliminates proteins that are potentially biologically interesting and detectable by multiple peptides if not at least two of these peptides meet all filter criteria. Automatic target list filtering with Skyline eliminated the vast majority of low-quality transitions. Moreover, utilizing all 24 samples from PS1 as a “training sample set” for several automatic steps and a final manual step of target list filtering further eliminated unreliable transitions with ragged peak shapes, interferences, inconsistent peak outlines or low signal-to-noise ratio.

The assay library permits large-scale (network-enabling) analyses of proteome dynamics in gills of threespine sticklebacks exposed to variable ecological contexts. A current limitation is that less than 10% of the overall proteome can be assayed with this assay library. Although it is unlikely that all proteins encoded by the genome are expressed in gills, even greater sensitivity and dynamic range of future instrumentation will gradually alleviate this limitation. The conceptual approach for assay library construction used in our study consisted of first acquiring high-quality fragment ion spectra in DDA mode, second, assigning peptide sequences to these spectra and annotating them to specific proteins, third, generating annotated spectral libraries, and fourth, compiling the final assay from the spectral library. This approach follows a recently described protocol (7, 8). However, here we use PEAKS Studio rather than the transproteomics pipeline (TPP) for consolidating multiple search engine results to annotate MSMS spectra from DDA data and Skyline rather than OpenSWATH (36) for spectral library construction and assay library filtering. Schubert and colleagues filtered out fragments smaller than 350 m/z or greater than 2,000 m/z as well as precursors that fall within the precursor isolation window from their assay library and they used a mass accuracy threshold within  $\pm 0.05$  m/z of the expected mass (8). We used similar criteria for the generation of the initial target list via the transition settings dialogue in Skyline. In step 1 of reducing the initial library to the final assay library the five most intense y and b ions were selected (min. fragment ion = ion 3, e.g. y4 even if y2 and y3 are more intense, max. fragment ion = last ion -1 excluding the precursor isolation window, m/z threshold  $\pm 0.035$  range) (Figure one a-d). Schubert and colleagues recommend that all assays in the final assay library have the same number of transitions to avoid mixed distributions that complicate statistical analysis. However, our inclusion of some precursors represented by four (rather than the nominal five) transitions in the assay library is accounted for since MSstats employs mixed linear models for statistical analysis (22). Removal of transitions that produce interferences in most or all training samples improves quantitative precision without impairing protein coverage of the assay library. In addition, four processed transitions still provide the highest separation power of the mProphet score (23).

### *Quantitative reproducibility of the stickleback gill assay library*

The reliability and quantitative reproducibility of the assay library was very high as assessed by analysis of three additional PSs, each consisting of 24 samples that were not included in the “sample training set” derived from PS1. No manual curation of the assay library was performed for any samples in PS2, PS3, and PS4 and all targeted data extraction procedures were done using Skyline automated workflows, including internal retention time alignment, mProphet peak scoring, and MSstats quantitative analysis (21-23). Visual inspection verified that transition peaks were properly assigned and outlined for precursors of proteins that were identified by MSstats analysis as significantly elevated or reduced in a particular population.

Since all raw data, quantitative results, and metadata are conveniently accessible in Panorama public (37) the corresponding Skyline files can be downloaded directly to reproduce all quantitative and statistical results. Although the population-specific abundance patterns of many proteins were very consistent in all four PSs, there were also differences for many proteins. Most of these differences were small. They can be attributed to three causes: First, individual differences in the expression of at least some proteins are expected between different fish. Second, the level of elevation or reduction of some proteins is slightly different, i.e. just above the significance threshold in some and slightly below that threshold in other PSs. Third, different population samplings for each of the four PSs may have pooled different subsets of the population. We view the last possibility as the least likely since all population samplings were performed under very similar conditions for a given population (same temperature, salinity, season, time of day, sampling method).

DIA datasets analyzed with the gill assay library represent a consistent matrix for powerful network modeling approaches, which can be complemented in the future by datasets for additional samples to enable comprehensive modeling of ecological and evolutionary effects on gill proteome dynamics. Although limited to a subset of the proteome (1506 proteins), the gill assay library allows robust and reproducible quantitation of a rich repertoire of striking population-specific proteome signatures. In

contrast to genomes that are stable within a single individual, the corresponding proteome is extremely dynamic depending on developmental, ecological, and cell/tissue contexts (38). Therefore, species- and tissue-specific proteomes can be considered excellent “reflections” of an organism’s life history exposures and experiences that provide information about its exposome (39, 40). Biological implications of the gill proteome signatures identified in this study are discussed below. An analysis of whether these proteome signatures represent the result of evolutionary/ genetic adaptations or physiological plasticity/ acclimation exceeds the scope of this study. Nevertheless, the identification of pertinent proteome signatures and the establishment of a gill assay library enables such future studies.

### ***Ecological significance of population-specific proteome signatures***

Pathway and GO enrichment approaches represent convenient tools for surveying broad biological implications of large datasets (41). We have utilized such approaches previously for threespine sticklebacks to interpret large proteomics datasets and condense the associated information (4, 5).

However, only known pathways can be analyzed pathway/GO enrichment approaches and representation of pathways and GO terms in existing databases is highly uneven. Protein annotation with GO terms and pathways is very heterogeneous ranging from none to hundreds per protein. Moreover, functional annotations are often based on only a handful of canonical model organisms studied in particular contexts. Thus, enrichment analyses are biased against poorly annotated pathways/ GO terms that may be more novel and unique in the particular tissue, species, and contexts studied. In addition, because the assay library covers less than 10% of the stickleback proteome the statistical power of enrichment analyses is lower compared to transcriptomics studies that utilize the entire genome for this purpose. Because of these limitations and caveats of such enrichment analyses we have complemented them with manual, literature-based assessment of pertinent functions of highly significant proteins.

### **Proteome signature of the Bodega Harbor population**

Marine sticklebacks from Bodega Harbor have the least unique gill proteome of all populations analyzed. A possible reason for this observation is the evolutionarily ancestral state of marine threespine stickleback populations (16), which may include most of the traits displayed by more specialized, derived populations at levels high enough to reflect a “species average”. Despite the lack of significant enrichment/ depletion of any PANTHER pathway/ GO term, a distinct proteome signature is evident even for this population. Although fewer in number compared to other populations, more than a hundred proteins are significantly elevated and reduced in BodHar fish. Elevated BodHar proteins cluster in KEGG pathways that are indicative of enhanced oxidative and lipid metabolism suggesting that energy utilization and/or storage in the form of lipids is elevated in the BodHar population. In addition, glutamine synthetase and antifreeze protein L12 promote osmotic homeostasis at elevated salinity by increasing the concentration of compatible osmolytes (42, 43). Reduced BodHar proteins cluster in growth-inhibitory (apoptosis, necroptosis), cell adhesion, extracellular pattern recognition, and ECM-related KEGG pathways. For instance, the marked depletion of transglutaminase 2 indicates reduced needs of the BodHar population for protein cross-linking (44), protection from apoptosis (45), ECM stabilization (46), and tissue damage repair (47). Overall, the BodHar gill proteome signature is consistent with the eco- and morpho-type of the BodHar population and suggests specific molecular targets that might underly enhanced growth and SW adaptation of BodHar fish.

### **Proteome signature of the LaBoRo population**

The LaBoRo population is unique among threespine sticklebacks since it chronically experiences temperatures of at least 30°C, which are lethal or highly detrimental for most other stickleback populations (48). Elevated LaBoRo proteins cluster in KEGG pathways that indicate enhanced roles of the extracellular matrix, cell adhesion, the plasma-membrane anchored cytoskeleton, and the complement system. Although many isoforms of common HSPs (HSP70, HSP60, HSP90) are included in the assay library none of them are elevated in the LaBoRo population. This unexpected finding may indicate that

these molecular chaperones are not limiting for proteostasis at chronically elevated temperature. However, two HSP47 (serpin 1) isoforms are strikingly elevated in gills of LaBoRo sticklebacks. HSP47 is a molecular chaperone for collagens and some other ECM proteins that promotes their maturation during thermal stress (49). Although collagens are not among the elevated ECM proteins in LaBoRo fish many other ECM and cell adhesion proteins are highly elevated suggesting a major role of extracellular matrix stabilization in LaBoRo stickleback gills. A key role of HSP47 for ECM reorganization has also been demonstrated in animal models of human diseases, including cardiac infarction (50), tumorigenesis (51, 52) and renal glomerulonephritis (53). Moreover, HSP47 plays critical roles for skeletal growth, ECM patterning, and fin regeneration in zebrafish (54). Elevated Hsp47 mRNA levels are correlated with increased temperature and heat-shock in many organisms (49, 55, 56), including fish (57), suggesting that the high habitat temperature causes severe ECM strain in gills of LaBoRo sticklebacks.

KEGG pathways that are most prominent among reduced LaBoRo proteins indicate a suppression of proteasome, lysosome, and ribosome functions as well as protein anabolism, adaptive immunity, and necroptosis. In addition, RNA/ DNA binding proteins and translation regulators/ factors are significantly enriched among reduced LaBoRo proteins supporting the notion of suppressed protein turnover in the LaBoRo population. A striking proteome feature of LaBoRo sticklebacks is the strong depletion of proteins involved in adaptive immunity while several complement proteins involved in innate immunity are elevated. Fish are ectotherms and thermal stress effects on the vertebrate complement system have been documented previously (58, 59). It should be pointed out that the fish gill is a highly vascularized tissue and that proteins detected in this study may be localized in branchial epithelial, vascular, and/ or extracellular compartments (60). The depletion of adaptive immunity proteins is exemplified by enormously reduced levels of interferon-induced GTP-binding protein Mx. This protein (Mx) is a very potent antiviral defense protein that protects cells from many RNA and DNA viruses (61, 62). Mx interacts with calcium signaling (63) and proteostasis pathways (e.g. the ER stress response) (64). Although causality between molecular LaBoRo proteome signatures and temperature is currently unclear, this population chronically experiences temperatures of at least 30°C, which are lethal or highly



detrimental for most other stickleback populations (48). In addition, strong temperature effects on the immune system have been reported for threespine sticklebacks (65) and other fish species (66). Whether caused by high habitat temperature or other environmental factors, the unique signature of immune system proteins in LaBoRo fish suggests that their adaptive immune functions are compromised. Induction of innate immunity via the complement system may represent a compensatory adaptive strategy. Alternatively, activation of the complement system may indicate increased levels of cellular and tissue-damage under the stressful habitat conditions experienced by LaBoRo sticklebacks. Such increased damage would exacerbate the demand for complement clearance of cellular debris (66, 67).

### **Proteome signature of the WesLag population**

The PANTHER integrin signaling pathway is significantly depleted from the elevated LaBoRo and WesLag proteins. This finding illustrates that this pathway is well represented by the 1506 assay library proteins but not elevated in either population. Significant enrichment of hydrolase and phosphatase activity among elevated WesLag proteins is exemplified by several highly elevated ATPases that are central for ion transport. The ribonucleoprotein complex is also enriched among elevated WesLag proteins, which cluster into calcium and GTPase signaling, actin cytoskeleton regulation, cell adhesion, pattern recognition, and protein translation KEGG pathways. The most highly elevated WesLag protein is chloride channel protein 2 (CLCN2). This ion channel and the elevated cation ATPases are regulated by environmental salinity in euryhaline fish, including sticklebacks (13, 60, 65, 68). We interpret these findings as evidence that WesLag fish experience frequent salinity fluctuations in their habitat, either because of tidal action or because of migrations between freshwater, brackish, and marine habitats. This anadromous population may have migrated just prior to collection from seawater to Westchester Lagoon to breed. Intracellular calcium, which controls many physiological processes of euryhaline fish including osmoregulation, is also regulated by the  $\text{Ca}^{2+}$ -ATPase (69). Besides calcium signaling, a prominent signal transduction feature of the WesLag-specific gill proteome signature are various GTPases, including several Ras-related small GTPases that are known to regulate structural stability of the cytoskeleton and

epithelial homeostasis and barrier function (70-72). These processes are key targets during cellular osmoregulation (73).

PANTHER enrichment of amino acid metabolism and blood coagulation among reduced WesLag proteins is accompanied by clustering of these proteins within endocannabinoid and PPAR signaling, spliceosome, lysosome, antigen processing and presentation, necroptosis, glutathione metabolism, and pentose phosphate KEGG pathways. Suppression of these processes in the WesLag population is supported by the functions of most significantly reduced proteins, for instance the marked depletion of proteins involved in immunity or RNA splicing. Reduced RNA processing and lower levels of complement and MHC proteins in this population indicate a lower demand for scavenging functions such as cleaning up and compensating cellular damage and pathogen debris by the complement system (67). Such functions may be more pronounced in the other populations if they experience greater levels of environmental stress and are more susceptible to pathogen invasion and cellular damage. Interestingly, the regulation of complement proteins and immunoglobulins is opposite in the WesLag population compared to the LaBoRo population, i.e. WesLag immunoglobulins are elevated while complement proteins are reduced. Whether temperature, specific pathogens, or other environmental factors are the cause for this proteome pattern remains to be determined.

### **Proteome signature of the LakSol population**

Elevated LakSol proteins are enriched for PANTHER hydrolase and peptidase activities and cluster predominantly in the KEGG pathways ribosome, lysosome, phagosome, antigen processing/ presentation, apoptosis and necroptosis. The most highly and significantly elevated LakSol proteins are caspase 1/interleukin 1 beta convertase, apoptosis-associated speck-like protein containing a CARD (PYCARD), cathepsins, and other proteases. The inflammasome comprises a multimeric protein complex consisting of caspase 1, PYCARD, and a nod-like receptor protein (NLRP, e.g. NLRP3) (74). Although not present in the assay library, the NLRP sensor of LakSol stickleback gill inflammasomes is likely NLRP3 because this sensor preferentially interacts with the PYCARD adaptor (75, 76). The PYCARD adaptor activates

the effector caspase 1 leading to proteolytic cleavage of an inactive cytokine precursor into active, pro-inflammatory interleukin-1 beta (74, 77-79). Inflammasome induction represents a response to parasitism in many species, including fishes (80-82). Freshwater sticklebacks are frequently parasitized (83) and we did observe cestode parasites (*Schistocephalus solidus*) in the body cavity of some LakSol sticklebacks during dissection, but never in any of the other populations. We excluded heavily parasitized fish during dissection but defense mechanisms against parasitism may be even more effective in fish from the same habitat that lack visible signs of parasitism. Potent inflammasome induction has also been reported in response to mercury exposure (84, 85). This finding is significant because Lake Solano is characterized by high mercury concentration resulting from historical gold mining activities in California (86). Intriguingly, cathepsins, which are also elevated in gills of LakSol sticklebacks, regulate the appearance and severity of mercury-induced inflammation (84). In addition, cathepsins are known to participate in inflammasome pathways in many other contexts (87-89).

Reduced LakSol proteins are depleted in PANTHER nucleic acid and RNA binding proteins suggesting that the levels of these proteins are maintained similar to those in other populations. Instead, reduced LakSol proteins cluster in focal adhesion, ECM-receptor interaction, tight junction, PI3K-Akt signaling, actin cytoskeleton regulation, and oxidative metabolism KEGG pathways. A reduction of oxidative metabolism is supported by a decrease in hemoglobin isoforms. Decreases in hemoglobin and erythrocytes have been observed as a result of parasitism in freshwater fish (90). For instance, the Brazilian cultivated sport fish *Leporinus microcephalus* has a reduced hematocrit, erythrocyte count, and hemoglobin concentration caused by the parasite *Goezia leporini* (91). Depletion of hemoglobins, iron transporting ferritins, and other plasma proteins may also be a consequence of parasitism in LakSol sticklebacks, but this hypothesis remains to be tested. As expected, proteins responsible for active NaCl secretion (e.g.  $\text{Na}^+/\text{K}^+/\text{2Cl}^-$  cotransporter) are reduced in LakSol fish indicating that their gills operate in ion absorption mode (60, 92) in their FW habitat.

## ***Conclusions and perspective***

To enable reliable and reproducible quantitation of a consistent set of proteins based on identical criteria in every sample we have developed an assay library for threespine stickleback gills. This resource empowers systematic large-scale analyses of the ecological context-dependence of gill proteome dynamics in threespine sticklebacks. The utility of this assay for ecological proteomics was demonstrated by comparing gill proteomes of four populations of threespine sticklebacks. Discrete and ecophysiologically meaningful proteome signatures of locally adapted populations have been identified. These signatures allow deep insight into mechanisms of biochemical adaptation and physiological plasticity in different environments. Future use of this assay library permits reliable and comparable assessment of the effects of any biological context on the gill proteome of threespine sticklebacks. For instance, it can be used in conjunction with controlled acclimation experiments that permit isolation of specific environmental variables such as salinity, temperature, and parasitism to correlate them with gill proteome signatures. Moreover, the gill assay represents a model for construction of assay libraries for other tissues. Multi-tissue assays can be used jointly to capture highly complex, ecologically relevant features of whole organism proteome dynamics. Moreover, a multi-tissue strategy combined with network modeling approaches will reveal environmental impacts on holistic scale proteome dynamics. Such standardized systems-level analyses promise to provide extensive insight into ecological and evolutionary impacts on complex vertebrates.

## **Acknowledgements**

We would like to thank Professor Wolfgang Polonik (UC Davis Department of Statistics) for providing guidance with the statistical analysis of data, Yafen Zhen and Tuan Pham (both UC Davis ANS Department) for help with the maintenance and care of sticklebacks and Dr. Alison Gardell (University of Alaska, Anchorage) for providing us with WesLag sticklebacks. This study was supported by NSF award 1656371 to DK.

## Data Availability

Raw mass spectrometry data, metadata, and processed data are available at the following public repositories:

MassIVE, ID MSV000081795 (<ftp://MSV000081795@massive.ucsd.edu>)

ProteomeXchange, ID PXD008395

(<http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD008395>)

CAMP proteome, AC CAMPDDA00023

([https://kueltzlab.ucdavis.edu/CAMP\\_dda\\_profiles.cfm?AC=CAMPDDA00023](https://kueltzlab.ucdavis.edu/CAMP_dda_profiles.cfm?AC=CAMPDDA00023))

Panorama Public, ID assigned when made public upon manuscript acceptance

(<https://panoramaweb.org/labkey/JoLi-01.url>)

CAMP proteome, AC CAMPDIA00001

([https://kueltzlab.ucdavis.edu/CAMP\\_dia\\_exp.cfm?ac=CAMPDIA00001](https://kueltzlab.ucdavis.edu/CAMP_dia_exp.cfm?ac=CAMPDIA00001))

## References

1. Marcon, E., Jain, H., Bhattacharya, A., Guo, H., Phanse, S., Pu, S., Byram, G., Collins, B. C., Dowdell, E., Fenner, M., Guo, X., Hutchinson, A., Kennedy, J. J., Krastins, B., Larsen, B., Lin, Z. Y., Lopez, M. F., Loppnau, P., Miersch, S., Nguyen, T., Olsen, J. B., Paduch, M., Ravichandran, M., Seitova, A., Vadali, G., Vogelsang, M. S., Whiteaker, J. R., Zhong, G., Zhong, N., Zhao, L., Aebersold, R., Arrowsmith, C. H., Emili, A., Frappier, L., Gingras, A. C., Gstaiger, M., Paulovich, A. G., Koide, S., Kossiakoff, A. A., Sidhu, S. S., Wodak, S. J., Graslund, S., Greenblatt, J. F., and Edwards, A. M. (2015) Assessment of a method to characterize antibody selectivity and specificity for use in immunoprecipitation. *Nat Methods* 12, 725-731
2. Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., and MacCoss, M. J. (2013) Multiplexed MS/MS for improved data-independent acquisition. *Nat Methods* 10, 744-746
3. Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol Cell Proteomics* 11, O111 016717

4. Kültz, D., Li, J., Paguio, D., Pham, T., Eidsaa, M., and Almaas, E. (2016) Population-specific renal proteomes of marine and freshwater three-spined sticklebacks. *J Proteomics* 135, 112-131
5. Kültz, D., Li, J., Zhang, X., Villarreal, F., Pham, T., and Paguio, D. (2015) Population-specific plasma proteomes of marine and freshwater three-spined sticklebacks (*Gasterosteus aculeatus*). *Proteomics* 15, 3980-3992
6. Heaven, M. R., Funk, A. J., Cobbs, A. L., Haffey, W. D., Norris, J. L., McCullumsmith, R. E., and Greis, K. D. (2016) Systematic evaluation of data-independent acquisition for sensitive and reproducible proteomics-a prototype design for a single injection assay. *J Mass Spectrom* 51, 1-11
7. Rost, H. L., Aebersold, R., and Schubert, O. T. (2017) Automated SWATH data analysis using targeted extraction of ion chromatograms. *Methods Mol Biol* 1550, 289-307
8. Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., Lam, H., Amodei, D., Mallick, P., MacLean, B., and Aebersold, R. (2015) Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat Protoc* 10, 426-441
9. Meyer, J. G., and Schilling, B. (2017) Clinical applications of quantitative proteomics using targeted and untargeted data-independent acquisition techniques. *Expert Rev Proteomics* 14, 419-429
10. Law, K. P., and Lim, Y. P. (2013) Recent advances in mass spectrometry: data independent analysis and hyper reaction monitoring. *Expert Rev Proteomics* 10, 551-566
11. Tinbergen, N., and Vaniersel, J. J. A. (1948) Displacement reactions in the three-spined stickleback. *Behaviour* 1, 56-63
12. Bell, M. A. F., Susan A. (1994) *The evolutionary biology of the threespine stickleback*, Oxford University Press, Oxford
13. Divino, J. N., Monette, M. Y., McCormick, S., Yancey, P. H., Flannery, K. G., Bell, M. A., Rollins, J. L., von Hippel, F. A., and Schultz, E. T. (2016) Osmoregulatory physiology and rapid evolution of salinity tolerance in threespine stickleback recently introduced to fresh water. *Evolutionary Ecology Research* 17, 179-201
14. Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., Brady, S. D., Zhang, H., Pollen, A. A., Howes, T., Amemiya, C., Lander, E. S., Di Palma, F., Lindblad-Toh, K., and Kingsley, D. M. (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55-61
15. Viitaniemi, H. M., and Leder, E. H. (2011) Sex-biased protein expression in threespine stickleback, *Gasterosteus aculeatus*. *Journal of Proteome Research* 10, 4033-4040
16. Wootton, R. J. (1976) *The biology of sticklebacks*, Academic Press, London

17. Ruiz-Campos, G., Camarena-Rosales, F., González-Acosta, A. F., Maeda-Martinez, A. M., García de León, F. J., Varela-Romero, A., and Andreu-Soler<sup>6</sup>, A. (2014) Estatus actual de conservación de seis especies de peces dulceacuícolas de la península de Baja California, México. *Revista Mexicana de Biodiversidad* 85, 1235-1248
18. Rainey, P. B., Remigi, P., Farr, A. D., and Lind, P. A. (2017) Darwin was right: where now for experimental evolution? *Curr Opin Genet Dev* 47, 102-109
19. Frazer, K. A., Elnitski, L., Church, D. M., Dubchak, I., and Hardison, R. C. (2003) Cross-species sequence comparisons: a review of methods and available resources. *Genome Research* 13, 1-12
20. Manzoni, C., Kia, D. A., Vandrovcova, J., Hardy, J., Wood, N. W., Lewis, P. A., and Ferrari, R. (2016) Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Briefings in bioinformatics*
21. Pino, L. K., Searle, B. C., Bollinger, J. G., Nunn, B., MacLean, B., and MacCoss, M. J. (2017) The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. *Mass Spectrom Rev* DOI: 10.1002/mas.21540, 1-16
22. Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., and Vitek, O. (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* 30, 2524-2526
23. Reiter, L., Rinner, O., Picotti, P., Huttenhain, R., Beck, M., Brusniak, M. Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat Methods* 8, 430-435
24. Benjamini, Y., and 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, 289-300
25. Kültz, D., Li, J., Gardell, A., and Sacchi, R. (2013) Quantitative molecular phenotyping of gill remodeling in a cichlid fish responding to salinity stress. *Mol Cell Proteomics* 12, 3962-3975
26. Wang, X., and Kültz, D. (2017) Osmolality/salinity-responsive enhancers (OSREs) control induction of osmoprotective genes in euryhaline fish. *Proc Natl Acad Sci U S A* 114, E2729-E2738
27. Abbatiello, S., Ackermann, B. L., Borchers, C., Bradshaw, R. A., Carr, S. A., Chalkley, R., Choi, M., Deutsch, E., Domon, B., Hoofnagle, A. N., Keshishian, H., Kuhn, E., Liebler, D. C., MacCoss, M., MacLean, B., Mani, D. R., Neubert, H., Smith, D., Vitek, O., and Zimmerman, L. (2017) New guidelines for publication of manuscripts describing development and application of targeted mass spectrometry measurements of peptides and proteins. *Mol Cell Proteomics* 16, 327-328
28. Sturn, A., Quackenbush, J., and Trajanoski, Z. (2002) Genesis: cluster analysis of microarray data. *Bioinformatics* 18, 207-208



29. Oliveros, J. C. (2015) Venny: An interactive tool for comparing lists with Venn's diagrams.  
<http://bioinfogp.cnb.csic.es/tools/venny/index.html>
30. Mi, H., Huang, X., Muruganujan, A., Tang, H., Mills, C., Kang, D., and Thomas, P. D. (2017) PANTHER version 11: expanded annotation data from gene ontology and reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 45, D183-D189
31. Koopmans, F., Ho, J. T. C., Smit, A. B., and Li, K. W. (2018) comparative analyses of data independent acquisition mass spectrometric approaches: DIA, WiSIM-DIA, and untargeted DIA. *Proteomics* 18
32. Holewinski, R. J., Parker, S. J., Matlock, A. D., Venkatraman, V., and Van Eyk, J. E. (2016) Methods for SWATH: data independent acquisition on tripleTOF mass spectrometers. *Methods Mol Biol* 1410, 265-279
33. Arnhard, K., Gottschall, A., Pitterl, F., and Oberacher, H. (2015) Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry. *Anal Bioanal Chem* 407, 405-414
34. Vidova, V., and Spacil, Z. (2017) A review on mass spectrometry-based quantitative proteomics: Targeted and data independent acquisition. *Anal Chim Acta* 964, 7-23
35. Bruderer, R., Sondermann, J., Tsou, C. C., Barrantes-Freer, A., Stadelmann, C., Nesvizhskii, A. I., Schmidt, M., Reiter, L., and Gomez-Varela, D. (2017) New targeted approaches for the quantification of data-independent acquisition mass spectrometry. *Proteomics* 17
36. Rost, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S. M., Schubert, O. T., Wolski, W., Collins, B. C., Malmstrom, J., Malmstrom, L., and Aebersold, R. (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* 32, 219-223
37. Sharma, V., Eckels, J., Taylor, G. K., Shulman, N. J., Stergachis, A. B., Joyner, S. A., Yan, P., Whiteaker, J. R., Halusa, G. N., Schilling, B., Gibson, B. W., Colangelo, C. M., Paulovich, A. G., Carr, S. A., Jaffe, J. D., MacCoss, M. J., and MacLean, B. (2014) Panorama: a targeted proteomics knowledge base. *J Proteome Res* 13, 4205-4210
38. Crowgey, E. L., Matlock, A., Venkatraman, V., Fert-Bober, J., and Van Eyk, J. E. (2017) Mapping biological networks from quantitative data-independent acquisition mass spectrometry: data to knowledge pipelines. *Methods Mol Biol* 1558, 395-413
39. Renz, H., Holt, P. G., Inouye, M., Logan, A. C., Prescott, S. L., and Sly, P. D. (2017) An exposome perspective: Early-life events and immune development in a changing world. *J Allergy Clin Immunol* 140, 24-40



40. Wild, C. P. (2005) Complementing the genome with an "exposome": the outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 14, 1847-1850
41. Yue, L., and Reisdorf, W. C. (2005) Pathway and ontology analysis: emerging approaches connecting transcriptome data and clinical endpoints. *Curr Mol Med* 5, 11-21
42. Tok, C. Y., Chew, S. F., Peh, W. Y., Loong, A. M., Wong, W. P., and Ip, Y. K. (2009) Glutamine accumulation and up-regulation of glutamine synthetase activity in the swamp eel, *Monopterus albus* (Zuiew), exposed to brackish water. *J Exp Biol* 212, 1248-1258
43. Scholander, P. F., van Dam, L., Kanwisher, J. W., Hammel, H. T., and Gordon, M. S. (1957) Supercooling and osmoregulation in arctic fish. *Journal of Cellular and Comparative Physiology* 49, 5-24
44. Collighan, R. J., and Griffin, M. (2009) Transglutaminase 2 cross-linking of matrix proteins: biological significance and medical applications. *Amino Acids* 36, 659-670
45. Fesus, L. (1998) Transglutaminase-catalyzed protein cross-linking in the molecular program of apoptosis and its relationship to neuronal processes. *Cell Mol Neurobiol* 18, 683-694
46. Grenard, P., Bresson-Hadni, S., El Alaoui, S., Chevallier, M., Vuitton, D. A., and Ricard-Blum, S. (2001) Transglutaminase-mediated cross-linking is involved in the stabilization of extracellular matrix in human liver fibrosis. *J Hepatol* 35, 367-375
47. Telci, D., and Griffin, M. (2006) Tissue transglutaminase (TG2): a wound response enzyme. *Front Biosci* 11, 867-882
48. Allen, J. R. M., and Wootton, R. J. (1982) The effect of ration and temperature on the growth of the three-spined stickleback, *Gasterosteus aculeatus* L. *Journal of Fish Biology* 20, 409-422
49. Ito, S., and Nagata, K. (2017) Biology of Hsp47 (SerpH1), a collagen-specific molecular chaperone. *Seminars in Cell & Developmental Biology* 62, 142-151
50. Hagiwara, S., Iwasaka, H., Shingu, C., Matumoto, S., Hasegawa, A., and Noguchi, T. (2011) Heat shock protein 47 (HSP47) antisense oligonucleotides reduce cardiac remodeling and improve cardiac function in a rat model of myocardial infarction. *The Thoracic and cardiovascular surgeon* 59, 386-392
51. Jiang, X., Zhou, T., Wang, Z., Qi, B., and Xia, H. (2017) HSP47 promotes glioblastoma stemlike cell survival by modulating tumor microenvironment extracellular matrix through TGF-beta pathway. *ACS Chem Neurosci* 8, 128-134
52. Zhu, J., Xiong, G., Fu, H., Evers, B. M., Zhou, B. P., and Xu, R. (2015) Chaperone Hsp47 drives malignant growth and invasion by modulating an ECM gene network. *Cancer Res* 75, 1580-1591

53. Kawazu, T., Nishino, T., Obata, Y., Furusu, A., Miyazaki, M., Abe, K., Koji, T., and Kohno, S. (2012) Production and degradation of extracellular matrix in reversible glomerular lesions in rat model of habu snake venom-induced glomerulonephritis. *Medical Molecular Morphology* 45, 190-198
54. Bhadra, J., and Iovine, M. K. (2015) Hsp47 mediates Cx43-dependent skeletal growth and patterning in the regenerating fin. *Mechanisms of Development* 138, 364-374
55. Ito, S., Ogawa, K., Takeuchi, K., Takagi, M., Yoshida, M., Hirokawa, T., Hirayama, S., Shin-ya, K., Shimada, I., Doi, T., Goshima, N., Natsume, T., and Nagata, K. (2017) A small-molecule compound inhibits a collagen-specific molecular chaperone and could represent a potential remedy for fibrosis. *Journal of Biological Chemistry*
56. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., and Hartl, F. U. (2013) Molecular chaperone functions in protein folding and proteostasis. *Annual Review of Biochemistry* 82, 323-355
57. Narum, S. R., and Campbell, N. R. (2015) Transcriptomic response to heat stress among ecologically divergent populations of redband trout. *BMC Genomics* 16, 103
58. Poskitt, T. R., and Poskitt, P. K. F. (1979) Temperature dependent activation of the alternate complement pathway by an IgG cryoglobulin. *American Journal of Hematology* 7, 147-154
59. Shah, T. A., Mauriello, C. T., Hair, P. S., Sandhu, A., Stolz, M. P., Bass, W. T., Krishna, N. K., and Cunnion, K. M. (2014) Clinical hypothermia temperatures increase complement activation and cell destruction via the classical pathway. *J Transl Med* 12, 181
60. Evans, D. H., Piermarini, P. M., and Choe, K. P. (2005) The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. *Physiol Rev* 85, 97-177
61. Fuchs, J., Holzer, M., Schilling, M., Patzina, C., Schoen, A., Hoenen, T., Zimmer, G., Marz, M., Weber, F., Muller, M. A., and Kochs, G. (2017) Evolution and antiviral specificities of interferon-induced Mx proteins of bats against ebola, influenza, and other RNA viruses. *J Virol* 91
62. Haller, O., Stertz, S., and Kochs, G. (2007) The Mx GTPase family of interferon-induced antiviral proteins. *Microbes Infect* 9, 1636-1643
63. Lussier, M. P., Cayouette, S., Lepage, P. K., Bernier, C. L., Francoeur, N., St-Hilaire, M., Pinard, M., and Boulay, G. (2005) MxA, a member of the dynamin superfamily, interacts with the ankyrin-like repeat domain of TRPC. *J Biol Chem* 280, 19393-19400
64. Numajiri Haruki, A., Naito, T., Nishie, T., Saito, S., and Nagata, K. (2011) Interferon-inducible antiviral protein MxA enhances cell death triggered by endoplasmic reticulum stress. *J Interferon Cytokine Res* 31, 847-856
65. Jürss, K., Bittorf, T., Vökler, T., and Wacke, R. (1982) Experimental studies on biochemical and physiological differences between the 3 morphs of the three-spined stickleback, *Gasterosteus aculeatus* L.

1. Gill Na/K-ATPase, muscle alanine aminotransferase and muscle aspartate aminotransferase activities. *Zool. Jahrb. Abt. Allg. Zool. Physiol. Tiere* 86, 267-272
66. Prohaszka, Z., Singh, M., Nagy, K., Kiss, E., Lakos, G., Duba, J., and Fust, G. (2002) Heat shock protein 70 is a potent activator of the human complement system. *Cell Stress Chaperones* 7, 17-22
67. Reddy, Y. N., Siedlecki, A. M., and Francis, J. M. (2017) Breaking down the complement system: a review and update on novel therapies. *Curr Opin Nephrol Hypertens* 26, 123-128
68. Zeng, J., Li, Z., Lui, E. Y., Lam, S. H., and Swaminathan, K. (2018) Tilapia and human CLIC2 structures are highly conserved. *Biochem Biophys Res Commun* 495, 1752-1757
69. Kültz, D. (2012) The combinatorial nature of osmosensing in fishes. *Physiology* 27, 259-275
70. Ikenouchi, J., and Aoki, K. (2017) Membrane bleb: A seesaw game of two small GTPases. *Small GTPases* 8, 85-89
71. Citalan-Madrid, A. F., Garcia-Ponce, A., Vargas-Robles, H., Betanzos, A., and Schnoor, M. (2013) Small GTPases of the Ras superfamily regulate intestinal epithelial homeostasis and barrier function via common and unique mechanisms. *Tissue Barriers* 1, e26938
72. Spiering, D., and Hodgson, L. (2011) Dynamics of the Rho-family small GTPases in actin regulation and motility. *Cell Adh Migr* 5, 170-180
73. Kültz, D., and Burg, M. B. (1998) Intracellular signaling in response to osmotic stress. *Contrib Nephrol* 123, 94-109
74. Strowig, T., Henao-Mejia, J., Elinav, E., and Flavell, R. (2012) Inflammasomes in health and disease. *Nature* 481, 278-286
75. Liu, S. B., Mi, W. L., and Wang, Y. Q. (2013) Research progress on the NLRP3 inflammasome and its role in the central nervous system. *Neurosci Bull* 29, 779-787
76. Ogura, Y., Sutterwala, F. S., and Flavell, R. A. (2006) The inflammasome: first line of the immune response to cell stress. *Cell* 126, 659-662
77. He, Y., Hara, H., and Núñez, G. (2016) Mechanism and regulation of NLRP3 inflammasome activation. *Trends in Biochemical Sciences* 41, 1012-1021
78. Guo, H., Callaway, J. B., and Ting, J. P. Y. (2015) Inflammasomes: mechanism of action, role in disease, and therapeutics. *Nature Medicine* 21, 677
79. Medzhitov, R. (2008) Origin and physiological roles of inflammation. *Nature* 454, 428-435
80. Lima-Junior, D. S., Mineo, T. W. P., Calich, V. L. G., and Zamboni, D. S. (2017) Dectin-1 activation during *Leishmania amazonensis* phagocytosis prompts syk-dependent reactive oxygen species production to trigger inflammasome assembly and restriction of parasite replication. *J Immunol* 199, 2055-2068

81. Ogryzko, N. V., Renshaw, S. A., and Wilson, H. L. (2014) The IL-1 family in fish: swimming through the muddy waters of inflammasome evolution. *Dev Comp Immunol* 46, 53-62
82. Angosto, D., Lopez-Castejon, G., Lopez-Munoz, A., Sepulcre, M. P., Arizcun, M., Meseguer, J., and Mulero, V. (2012) Evolution of inflammasome functions in vertebrates: Inflammasome and caspase-1 trigger fish macrophage cell death but are dispensable for the processing of IL-1beta. *Innate Immun* 18, 815-824
83. Nishimura, N., Heins, D. C., Andersen, R. O., Barber, I., and Cresko, W. A. (2011) Distinct lineages of *Schistocephalus* parasites in threespine and ninespine stickleback hosts revealed by DNA sequence analysis. *PLoS ONE* 6, e22505
84. Toomey, C. B., Cauvi, D. M., Hamel, J. C., Ramirez, A. E., and Pollard, K. M. (2014) Cathepsin B regulates the appearance and severity of mercury-induced inflammation and autoimmunity. *Toxicol Sci* 142, 339-349
85. Park, E. K., Mak, S. K., Kültz, D., and Hammock, B. D. (2007) Evaluation of cytotoxicity attributed to thimerosal on murine and human kidney cells. *J Toxicol Environ Health A* 70, 2092-2095
86. Sparks, G. C. (2016) Mercury and methylmercury related to historical mercury mining in three tributaries to Lake Berryessa, Upper Putah Creek watershed, California. *MS thesis (Geology)*, p. 278, California State University, Sacramento, Sacramento
87. Bai, H., Yang, B., Yu, W., Xiao, Y., Yu, D., and Zhang, Q. (2018) Cathepsin B links oxidative stress to the activation of NLRP3 inflammasome. *Exp Cell Res* 362, 180-187
88. Wang, L., Chen, Y., Li, X., Zhang, Y., Gulbins, E., and Zhang, Y. (2016) Enhancement of endothelial permeability by free fatty acid through lysosomal cathepsin B-mediated Nlrp3 inflammasome activation. *Oncotarget* 7, 73229-73241
89. Bruchard, M., Mignot, G., Derangere, V., Chalmin, F., Chevriaux, A., Vegran, F., Boireau, W., Simon, B., Ryffel, B., Connat, J. L., Kanellopoulos, J., Martin, F., Rebe, C., Apetoh, L., and Ghiringhelli, F. (2013) Chemotherapy-triggered cathepsin B release in myeloid-derived suppressor cells activates the Nlrp3 inflammasome and promotes tumor growth. *Nat Med* 19, 57-64
90. Shahi, N., Yousuf, A. R., Rather, M. I., Ahmad, F., and Yaseen, T. (2013) First report of blood parasites in fishes from Kashmir and their effect on the haematological profile. *Open Veterinary Journal* 3, 89-95
91. Martins, M. L., Tavares-Dias, M., Fujimoto, R. Y., Onaka, E. M., and Nomura, D. T. (2004) Haematological alterations of *Leporinus macrocephalus* (Osteichthyes: Anostomidae) naturally infected by *Goezia leporini* (Nematoda: Anisakidae) in fish pond. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia* 56, 640-646

92. Kültz, D. (2015) Physiological mechanisms used by fish to cope with salinity stress. *J Exp Biol* 218, 1907-1914

## Figure legends

**Figure one:** Mass accuracy and retention time reproducibility for all samples analyzed in this study.

The mass error (ppm) for all transitions present in the final assay library is shown in the top row for PS1 (**A**), PS2 (**B**), PS3(**C**), and PS4 (**D**). The corresponding retention time reproducibility of the iRT standards is shown in the bottom row. Figures were generated with Skyline 3.6 software (MacCoss Lab., University of Washington).

**Figure two:** Composition of the DIA target lists before and after filtration. The numbers of proteins (**A**), peptides (**B**), precursors (**C**), and transitions (**D**) in the initial (white bars labeled SL for Spectral Library) and final (black bars labeled AL for assay library) target lists is depicted. Library filtration steps two to six are explained in the text. **E**) Frequency of proteins represented by two to ten peptides in the final assay library. **F**) Frequency distributions of fragment ion types represented in the final assay library. The data were generated with Skyline 3.6 (MacCoss Lab., University of Washington).

**Figure three:** **A)** Fold change (FC) and coefficient of variation (CV) depending on number of biological replicates at a statistical power of 0.8 and false discovery rate (FDR) of 1%. **B)** Volcano plots for each population sampling (PS) and each population (Laguna de la Bocana del Rosario = LaBoRo, Lake Solano = LakSol, Westchester Lagoon = WesLag, Bodega Harbor = BodHar). For clarity, only proteins that are significant in at least one PS and fall within FC -4 (0.25) to +4 and  $-\log_{10} p$  between 0 and 4 are plotted. For complete coverage refer to table S4. Blue circles represent proteins that are significantly reduced in all PSs at  $p < 0.01$  and  $FC \geq 2$ , blue triangles represent proteins that are significantly reduced at  $p < 0.05$  and  $FC \geq 2$  but not at  $p < 0.01$ , blue squares represent proteins that are not significantly reduced for the corresponding PS. Orange circles represent significantly elevated proteins at  $p < 0.01$ , red triangles represent significantly elevated proteins at  $p < 0.05$  and  $FC \geq 2$  but not at  $p < 0.01$ , yellow squares represent proteins that are not significantly elevated for the corresponding PS.

**C)** Number of false positive proteins at  $p < 0.05$  and  $p < 0.01$  (cumulative bars) for PS1 to PS4 (from left to right) and each population. Missing bars indicate zero false positives. False positives are expressed as the number of proteins exceeding the significance threshold in the opposite FC direction for the PS depicted compared to any of the other three PSs (e.g. FC = 2.0 in PS1 but 0.5 in PS2). **D)** Number of false negative proteins at  $p < 0.05$  and  $p < 0.01$  (cumulative bars) for PS1 to PS4 (from left to right) and each population. False negatives are expressed as the number of proteins exceeding the FC threshold of 2.0 but not the statistical significance threshold.

**Figure four:** Proteins from threespine stickleback gills that are significantly elevated and reduced in each population relative to the other populations. Data from all population samplings were combined for this analysis. The top row shows the volcano plots with significant proteins depicted in blue (reduced) and orange (elevated). Underneath are the corresponding heat maps sorted by first elevated and then reduced proteins. Each heat map depicts data for 96 biological replicates from left to right as follows: 24 LaBoRo, 24 LakSol, 24 WesLag, 24 BodHar. The color scale of the heat map ranges from -3 (threefold reduction, blue) to +3 (threefold increase, orange). Note that the number of significant proteins is different for each population and that heat maps are stretched to fit the panel. Miniature labels designate samples at the top and protein accession number on the right of each heat map. They can be visualized by enlarging the PDF file. These data are also available in numerical form in table S5.

**Figure five:** Numbers of significantly elevated (left panels) and reduced (right panels) proteins that cluster in KEGG pathways (minimum = 5, grey bars) and are significantly enriched (black bars) or depleted (white bars) in PANTHER pathways and gene ontology (GO) categories. Data are shown for the Laguna de la Bocana del Rosario (**A, B**), Lake Solano (**C, D**), Bodega Harbor (**E, F**), and Westchester Lagoon (**G, H**) populations.

**Figure six:** Proteins from threespine stickleback gills with the greatest and most significant abundance differences in the Laguna de la Bocana del Rosario (LaBoRo) population compared to the other populations. Data for significantly elevated (**A**) and reduced (**B**) proteins represent the sum of the means of all four population comparisons (PSs). Error bars represent the SEM of the cumulative dataset from all four PSs (n = 24 per population). For each protein the reproducibility of the overall trend can be assessed by inspecting the contributions of each PS, which are indicated by different black and white shading intensity.

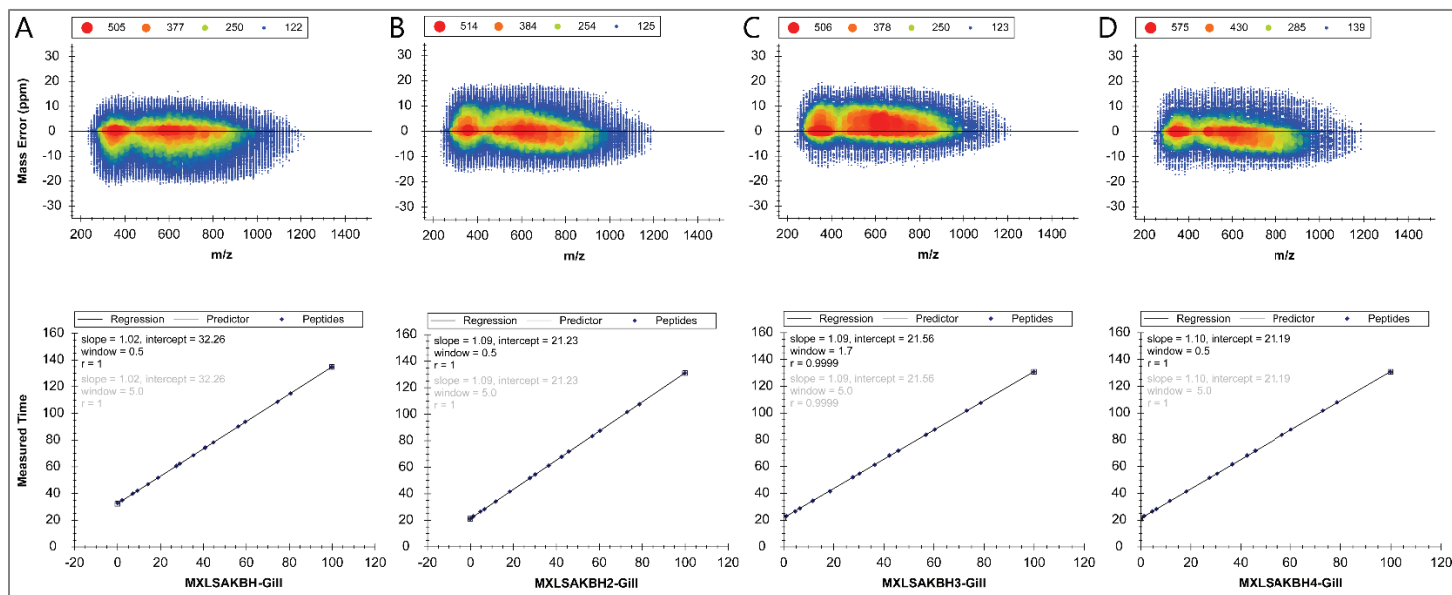
**Figure seven:** Proteins from threespine stickleback gills with the greatest and most significant abundance differences in the Lake Solano (LakSol, **A and B**) and Bodega Harbor (BodHar, **C and D**) populations compared to the other populations. Data for significantly elevated (**A, C**) and reduced (**B, D**) proteins represent the sum of the means of all four population comparisons (PS). Error bars represent the SEM of the cumulative dataset from all four PSs (n = 24 per population). For each protein the reproducibility of the overall trend can be assessed by inspecting the contributions of each PS, which are indicated by different black and white shading intensity.

**Figure eight:** Proteins from threespine stickleback gills with the greatest and most significant abundance differences in the Westchester Lagoon (WesLag) population compared to the other populations. Data for significantly elevated (**A**) and reduced (**B**) proteins represent the sum of the means of all four population comparisons (PSs). Error bars represent the SEM of the cumulative dataset from all four PSs (n = 24 per population). For each protein the reproducibility of the overall trend can be assessed by inspecting the contributions of each PS, which are indicated by different black and white shading intensity.

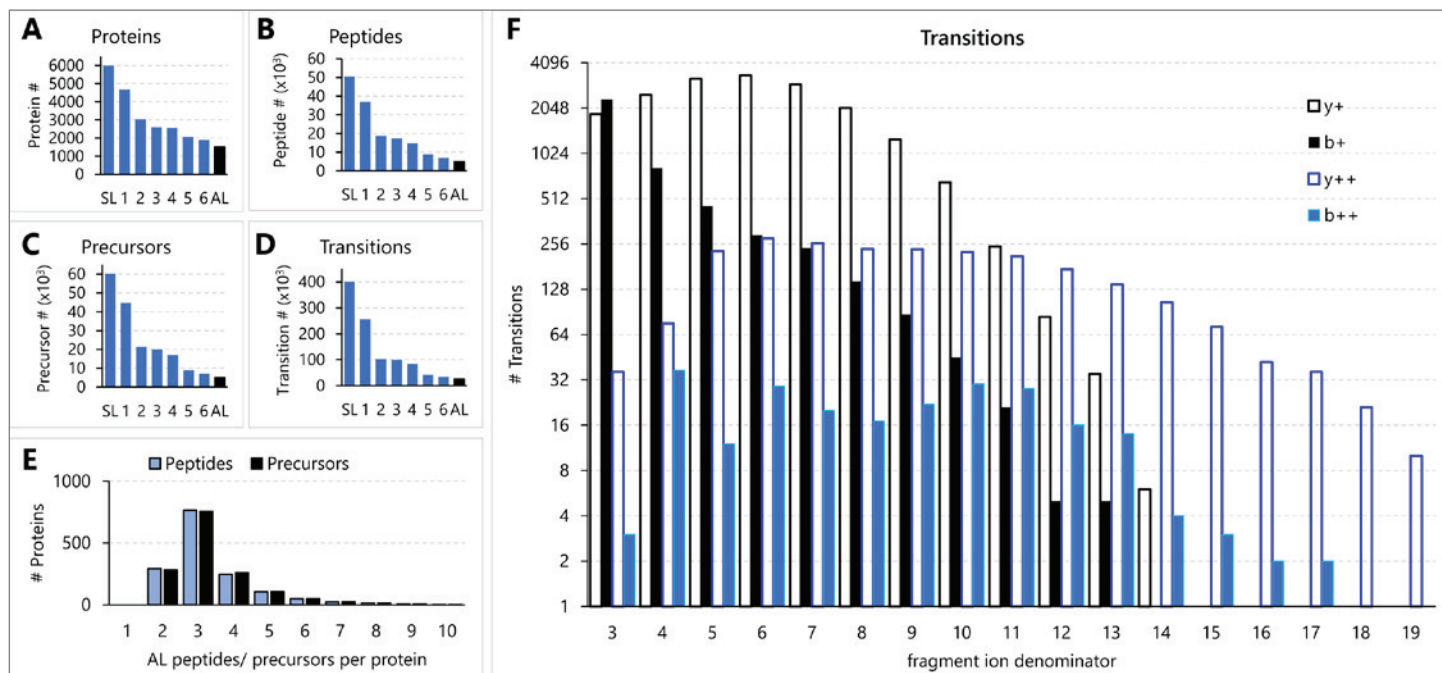


## Figures

*Figure one*



*Figure two*



**Figure three**

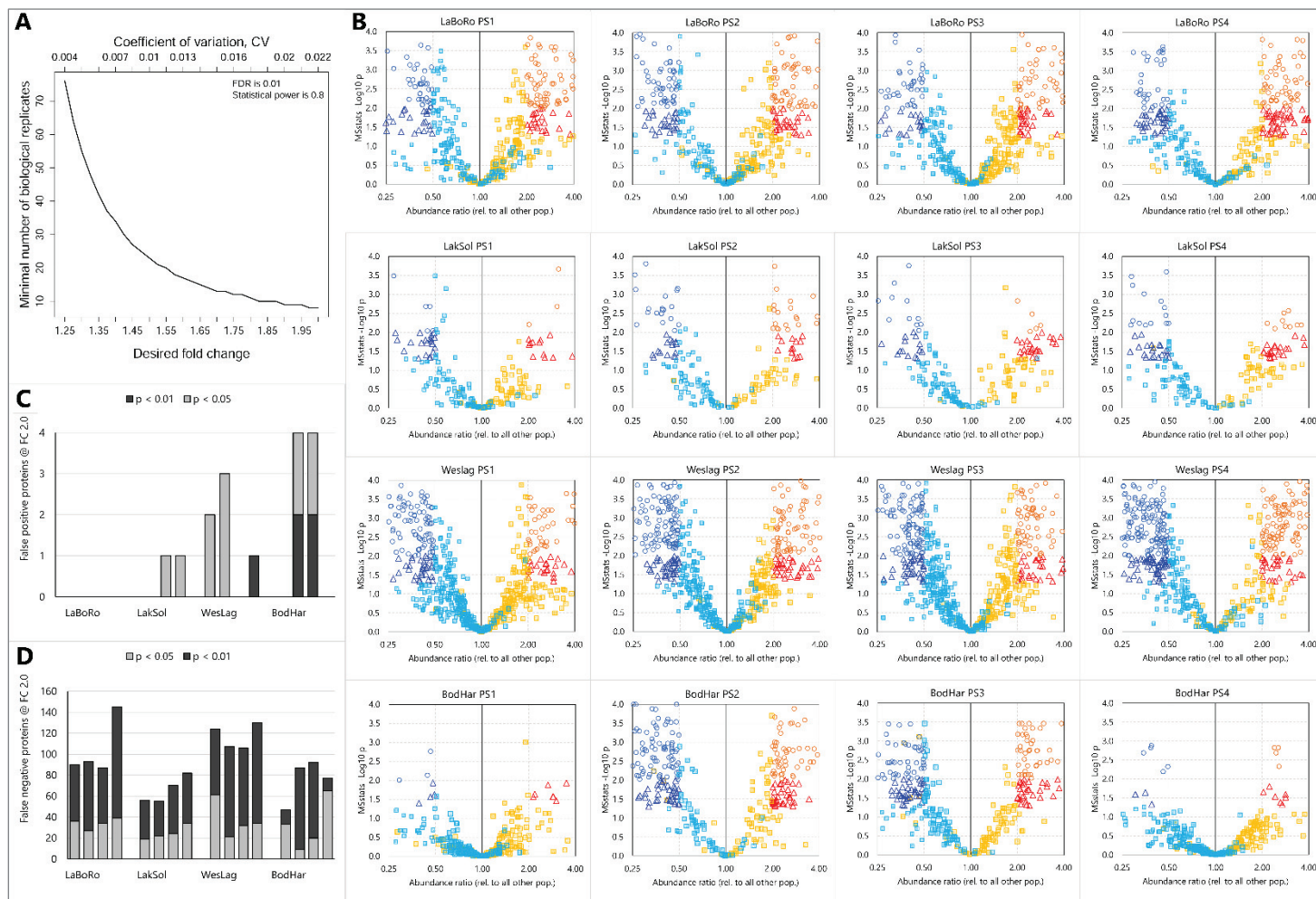
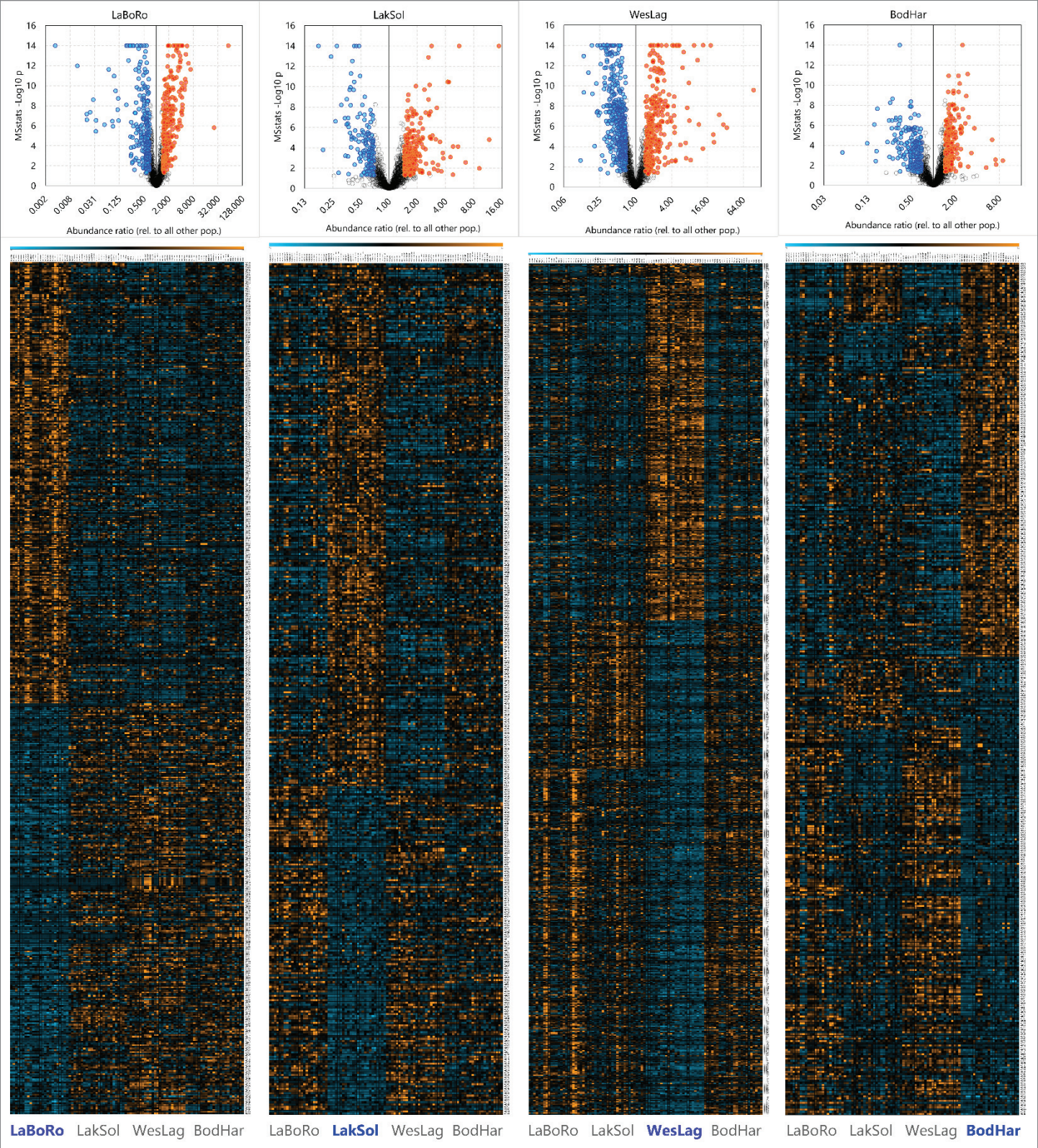


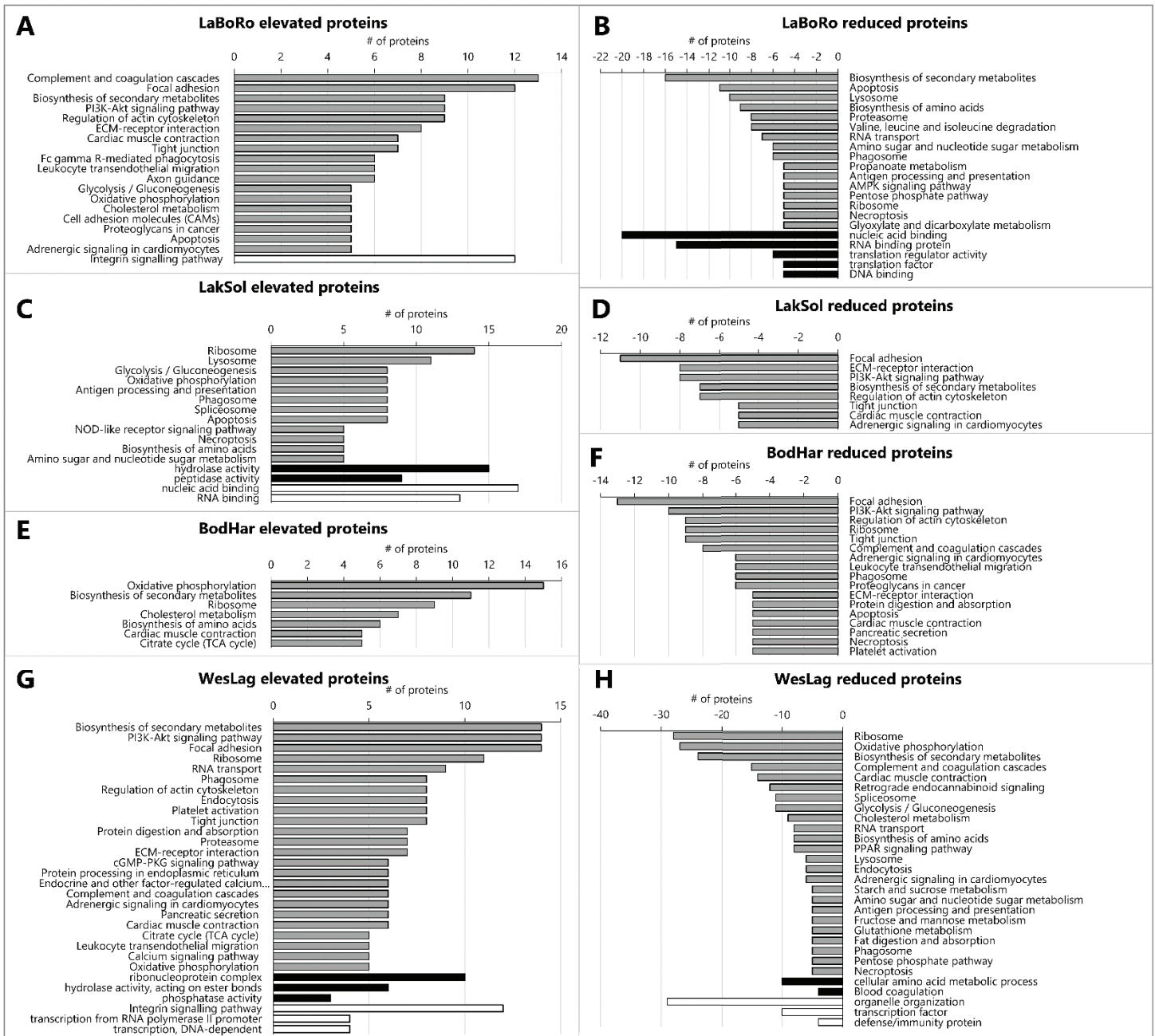


Figure four

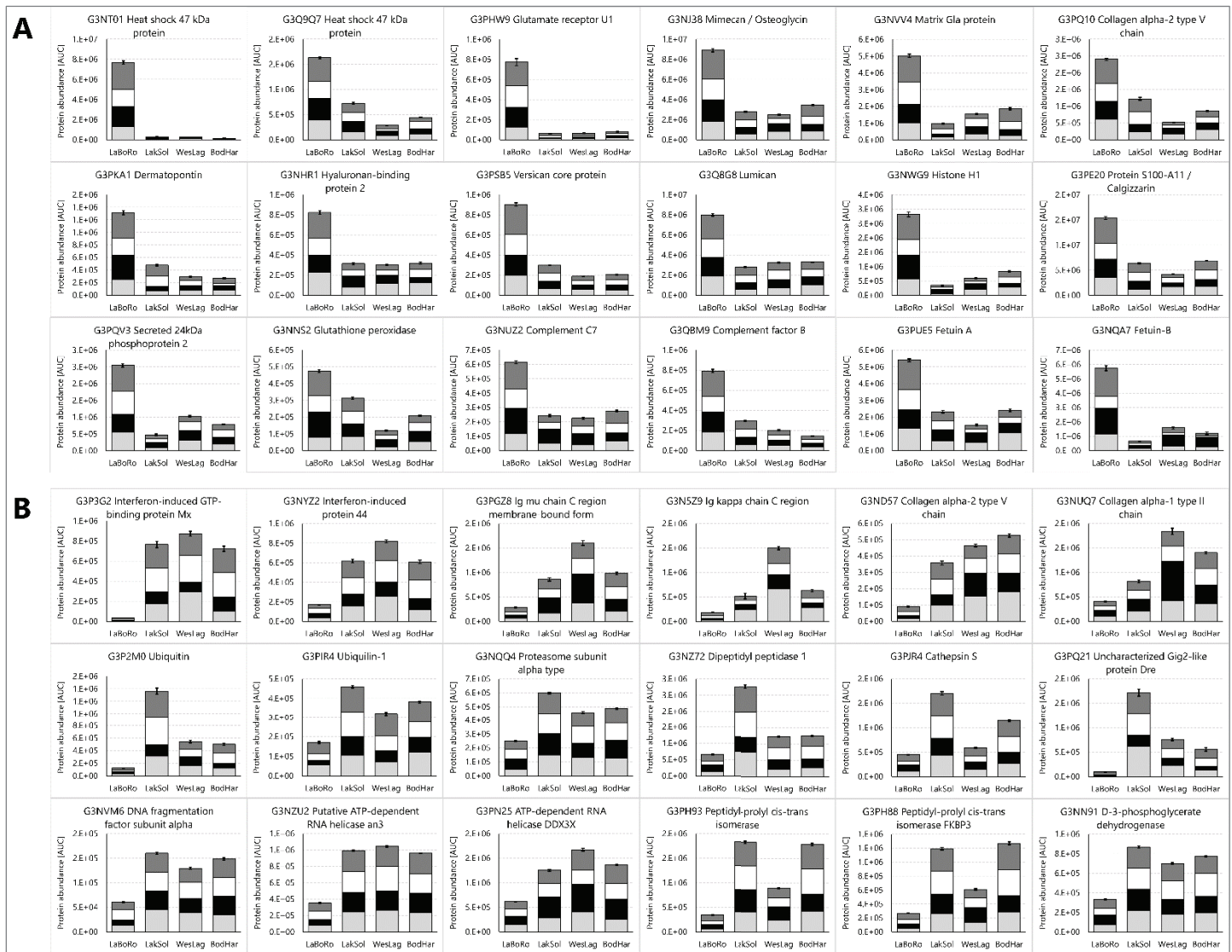




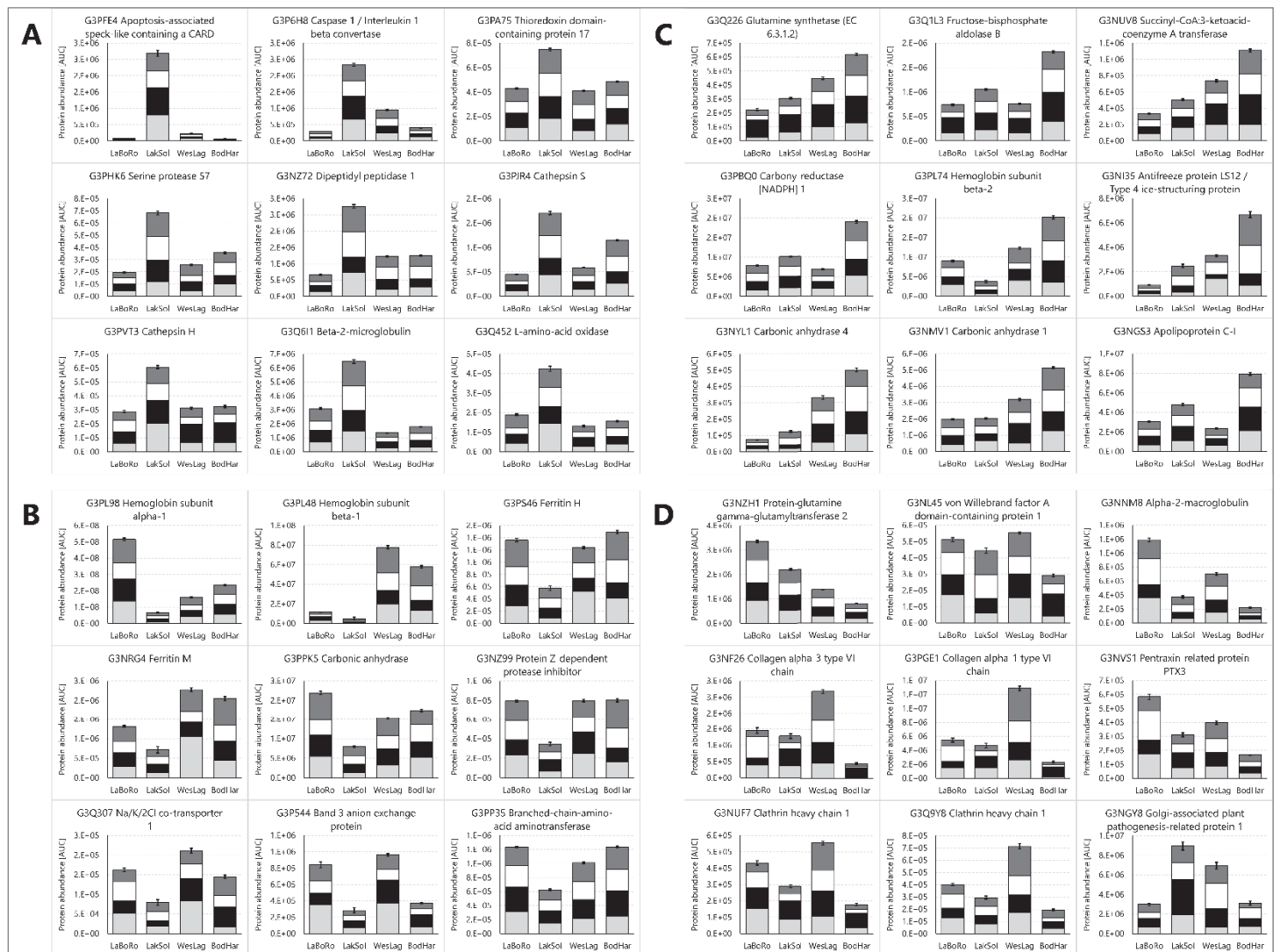
**Figure five:**



**Figure six:**



**Figure seven**



**Figure eight**

