Comisión Nacional de Investigación Científica y Tecnológica – CONICYT

FONDAP Centers of research Program

Annual Progress Report - 2017 (Year 7)

Guidelines:

The report should be written following the format specified hereafter. Both a printed (report and excel spreadsheets) and an electronic version must be sent to the following address:

PROGRAMA CENTROS DE EXCELENCIA FONDAP
CONICYT
Moneda 1375, Floor 9
Santiago

E-mail: mcamelio@conicyt.cl
emedina@conicyt.cl
Phone: (56) 224 354 327

For future inquiries, please contact:

María Eugenia Camelio
FONDAP Program Director
E-mail: mcamelio@conicyt.cl

I. PRESENTATION

PERIOD REPORTED: 6th Year [ ] 7th Year [X] 8th Year [ ] 9th Year [ ] 10th Year [ ]

PERIOD COVERED: From: January 2017 To: December 2017

NAME OF THE CENTER **Center for Genome Regulation**		**CODE** 15 09 00 07
DIRECTOR OF THE CENTER Miguel L Allende	**E-MAIL** mallende@uchile.cl	**SIGNATURE**
DEPUTY DIRECTOR Martín Montecino L.	**E-MAIL** mmontecino@unab.cl	**SIGNATURE**
SPONSORING INSTITUTION Universidad de Chile		
ASSOCIATED INSTITUTION(S) (if applicable) Universidad Andrés Bello; Pontificia Universidad Católica de Chile		
CENTER WEBSITE ADDRESS www.genomacrg.cl		

DATE: 31/01/18

Research Lines

N°	Research Line	Objective	Principal Researcher	Associated Researcher(s)
1	Aim 1	A phylogenomic and systems biology approach to identify genes underlying plant survival in marginal soils.	Rodrigo Gutiérrez	Fernán Federici
2	Aim 2	Metagenome of the altiplano soils: plant-microbiome interaction.	Mauricio González	Verónica Cambiazo
3	Aim 3	Regulatory landscape plasticity as an evolutionary driver in the genomes of Cyprinidontiform fish.	Martín Montecino Miguel Allende	Christian Hodar
4	Aim 4	Identification of genomic signatures defining metabolic networks that provide unique features to cope with environmental stresses in plants.	Ariel Orellana	Claudio Meneses
5	Aim 5	Gene expression control and regulatory networking.	Alejandro Maass	Alvaro Glavic

II. EXECUTIVE SUMMARY

The CGR is seven years old as of this report and its consolidation as a leading institution in Chilean science has been the main outcome of its maturation. As has been customary, our scientific output is steady as we continue to show a high degree of impact by all measurement standards. Our young scientists have successfully incorporated into the research field, in academia or industry, and the number of trainees remains high. The opportunities for international exchanges and networking are ever present and have allowed for numerous new interactions with the wider community. We have also established links with diverse partners in the private sector, increasing our role in the applications derived from our research. This report will reflect the scientific advances achieved but also the vision we are developing for continuity of the CGR, as we are firmly committed to formulating long term goals and a plan that will move us forward for another 10 years.

All of the research lines that we are engaged in show progress and results that have been building continuously towards a new understanding of genomic and evolutionary processes. The principal aim of characterizing the genomic strategies used by organisms subjected to extreme conditions has proven to be challenging, but also, rewarding. The analysis of gene network evolution in desert plants has revealed convergent pathways in phylogenetically distant species, revealing how they cope with dramatically low levels of nutrients and water. Part of the answer has also arisen from our analysis of the microbial communities that coexist with the desert plants. The structure of these populations of microbes is highly specific to the rhizosphere and reveals mutual dependence. We have been sampling the desert for five consecutive years and will continue to do so as the appearance of certain species is critically dependent on the environmental conditions, which are highly variable. Thus, our analysis involves not only a comprehensive survey of almost all desert life but it provides a representation of its temporal fluctuations, an important consideration in the face of climate change. Not far away from our transect where plants and microbes are collected, we find *Orestias* fish, unlikely inhabitants of small geothermal-associated streamlets in high altitude salt pans. We have finalized the analysis of its genome and will be reporting on the features of genes that allow it to cope with such a harsh environment. Likewise, annual fish of the Atlantic coast, where fish have to deal with ephemeral ponds by surviving a dry season, have also been the subject of our genomic analysis. Together with the identification of selective traits that confer them with striking survival strategies, they show an incredibly rapid evolution of their genome structure, suggesting that phenotypic plasticity may arise from parallel flux and variation at the genomic level. Finally, the work on transcriptomics of the Andean toad, *Rhinella spinulosa*, was published this year, a paper that has as authors all six of the CGR PIs. Together with our work on the genomes of natural populations of organisms, we have continued to use our expertise to describe and analyze the genomes of species that are important to the Chilean economy. Our group had a crucial role in the elucidation of the extremely complex Atlantic Salmon genome, we are following this up with work in other salmonids, and in reporting on genomes of diverse strains of its pathogenic bacteria, *Piscirickettsia salmonis*. Likewise, we continue to generate data on diverse fruit crops that are agricultural and economic staples such as grape and peach as well as on other commercially important species. Our work on molecular and cellular

pathways and on epigenomic mechanisms continued to be the main research output of the Center in terms of papers, theses and congress presentations.

Our training efforts have maintained a numerous cohort of young scientists at the formative stage. Students have had the opportunity to travel to congresses and to carry out internships at centers and institutes worldwide, expanding their horizons and opening opportunities for further training abroad. These experiences have benefitted from the network of collaborative interactions we have built over the years. Some of the most relevant during 2017 were those with Japanese, Australian, European and North American centers of excellence. Within the region, close ties were maintained or established with colleagues in Uruguay, Colombia and Brazil. Numerous students from those and other countries visited the CGR labs, enriching their scientific careers.

Outreach is an integral part of our efforts as a center. We appeared consistently in Chilean media, print, radio and web-based outlets. However, we highlight the work we have done in schools, training teachers and hosting students. We have developed portable lab modules that visit schools and we have organized project based activities that have been very successful. We have already begun to organize a nation-wide activity that will launch the 1000 Genomes project. It will be a sequencing project carried out by school children using a simple technology developed for field work in genomics and we have the support of four Chilean centers of excellence who will be co-sponsors of the project. We have also had expressions of interest from the private sector, which may be a sign that we can break the barrier that has precluded investment in Chile's R&D by this player. Importantly, we envision this initiative as the path forward for our and other centers, as we must combine scientific excellence with societal benefit as the drivers of our research.

The CGR has functioned without any significant hindrance during 2017. As reported previously, the Directorship switched between Dr. Allende and Dr. González for one year as the former carried out a sabbatical stay; the organizational structure reverted back to its original form in July 2017. We have had no issues with our financial reporting and continue to be congratulated for our timely and flawless account rendering, a tribute to our staff. We maintain our commitment to hold weekly PI meetings and our trademark "jamborees", brainstorming and progress reporting sessions where all members of a working group come together. We also had a two-day long CGR retreat, where all labs had a chance to present their work and receive feedback; this event was in direct response to a suggestion from our reviewers.

Finally, we take this opportunity to comment on the current state of Chilean science. The imminent change of government comes without a clear definition on the future of policy or support structure for Chile's scientific community. While legislation that would create a Ministry of Science has passed one of the chambers, it is unclear whether it will be a solution to the lack of planning and coherence of the diverse initiatives that are currently available. It also does not guarantee a stable environment for long-term funding, something essential to centers like ours, which require years to develop and where results are forthcoming in a timeframe measured in decades. We will have to dedicate part of our time in the final three years of funding to explore ways in which to sustain this research effort and the strong investment in human capital we have generated.

III. ADMINISTRATIVE ASPECTS

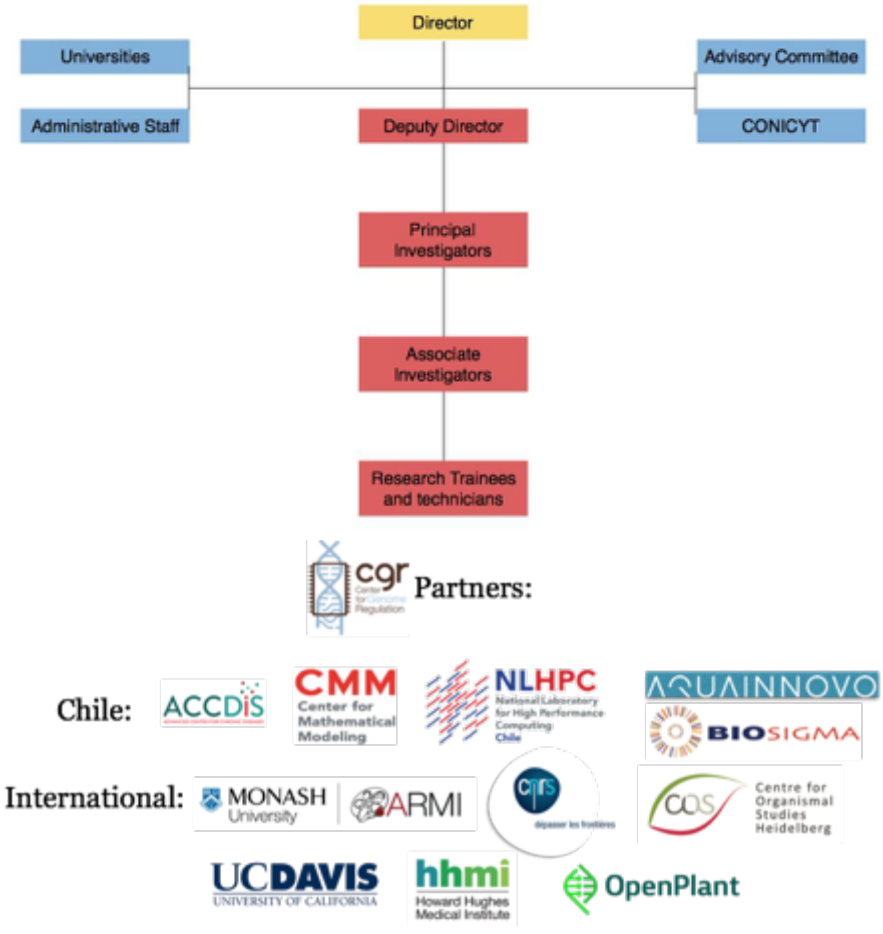
1. **Budget execution**: Describe and justify any budgetary modifications (itemized) of the original proposal.

   None of significance.

2. **Accomplishment of institutional commitments:** describe any difficulty(ies) encountered regarding this aspect.

Our academic institutions have complied with their mandate to support the Center.

3. **Organizational Chart:** Present an organizational chart of the Center depicting its main links to companies, associated institutions, and other units within the same institution.



4. **Personnel:** Provide a table indicating all personnel involved in the operation of the Center during the reported period, including names, position within the center (e.g. associate researcher, post doc, student, technician, etc.) and the number of hours committed to the Center.
   In addition, in no more than one page, provide a brief academic biography for each new researcher recruited by the Center.

Position	No.	Name	Hours
Director	1	Miguel L Allende	44
Deputy Director	2	Martín Montecino	26
Principal Investigator	3	Alejandro Maass	26
	4	Mauricio González	26
	5	Rodrigo Gutiérrez	26
	6	Ariel Orellana	26
Associate Investigator	7	Verónica Cambiazo	8
	8	Alvaro Glavic	8
	9	Christian Hodar	8
	10	Claudio Meneses	8
	11	Fernán Federici	8
Adjunct Investigator	12	Claudio Latorre	0
	13	Marco Mendez	0
	14	Mauricio Latorre	0
Postdoctoral Fellow	1	Francisca Diaz	44
	2	Jose Alvarez	44
	3	Viviana Araus	44
	4	Soledad Undurraga	44
	5	Orlando Contreras	44
	6	Grace Armijo	44
	7	Gino Nardocci	44

	8	Mauricio Saez Venegas	44
	9	Raquel Marina	44
	10	Laura Guajardo	44
	11	Hugo Sepulveda	44
	12	Mario Sanchez	44
	13	Ángela Cintolesi	44
	14	Luisa Pereiro	44
	15	Diego Rojas	44
	16	Myra Chávez	44
	17	Constanza Vásquez	44
	18	Matías Medina	44
	19	Gonzalo Olivares	44
	20	Esteban Contreras	44
	21	Fernanda Lisbona	44
	22	Adrian Moreno Vilches	44
	23	Ricardo Nilo	44
	24	Paula Vizoso	44
	25	Scarleth Bravo	44
	26	Jonathan Maldonado	44
	27	Nicolas Tobar	44
	28	Beatriz Fernandez	44
Students	1	Salomé de las Nieves Muñoz Sánchez	44
	2	Carlos Muñoz Montecinos	44
	3	Jorge Zúñiga	44
	4	Geraldine Del Carmen Aedo Vielma	44
	5	Gabriela Carrasco	44
	6	Isaak Núñez	44

	7	Tamara Matute	44
	8	Ma Paz Cortés	44
	9	Alex Di Genova	44
	10	Martín Rios	44
	11	Sandra Edwards Jorquera	44
	12	Nicolás Cumplido Salas	44
	13	Rodrigo Alonso Morales Castro	44
	14	Javiera Fernanda De la Paz Montt	44
	15	Claudia Molina Pelayo	44
	16	Andrea Constanza González Aguilar	44
	17	Kareb Inger Fehrmann Cartes	44
	18	Felipe Gajardo	44
	19	Fernanda Lourido Araneda	44
	20	Cristina Andrea Muñoz Rehbein	44
	21	Cristina Andrea Muñoz Rehbein	44
	22	Emiliano Molina	44
	23	Emiliano Molina	44
	24	Isabel Fredes	44
	25	Tomas Moyano	44
	26	Sebastián Moreno	44
	27	Catalina Ibarra	44
	28	Ariel Cerda	44
	29	María Paz Medina	44
	30	Valentina Zapata	44
	31	Alejandro Fonseca	44
	32	Susan Hitschfelt	44
	33	Macarena Muñoz	44

	34	Margarita Carrasco	44
	35	Sebastian Abarzua	44
	36	Esteban Quezada	44
	37	Dante Travisany	44
	38	Anibal Riveros Orellana	44
	39	Alonso Espinoza Rojas	44
	40	René Gonzalo Poblete Orrego	44
	41	Sergio González Itier	44
	42	Deborah Cuper	44
	43	Pamela Aravena	44
	44	Javiera Ortiz.	44
	45	Khantati Hauyon	44

5. **Changes in research personnel:** Describe any changes in the principal and associate researchers relative to the original project.

   No changes.

6. **Advisory committee:** describe its tasks, the frequency of meetings, and usefulness of the advice provided to the Center. Also, report on the availability of the committee to assist the Center.

While we have maintained communication with our Advisory panel of three renowned scientists, we have not been able to find a convenient date for a face to face meeting with our Center's investigators. In part, this was due to the sabbatical year of our Director. We have agreed on a tentative date of April 2018 to resume our meetings with the panel.

IV. OBJECTIVES AND RESULTS ATTAINED (Maximum 20 pages)

1. RESULTS OBTAINED RELATIVE TO CENTER OBJECTIVES

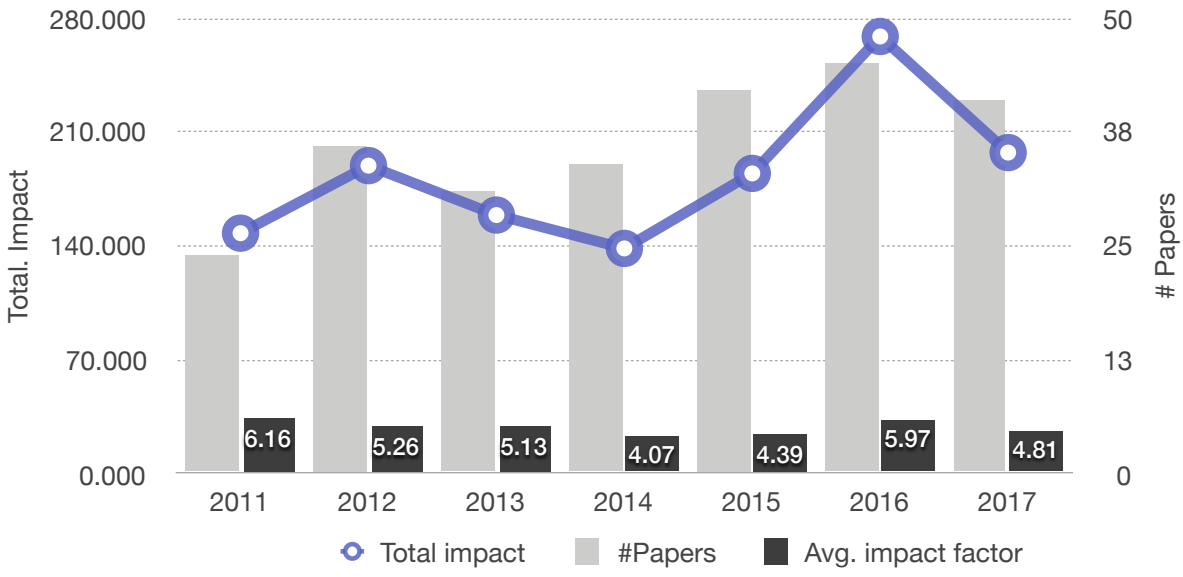
a. Considering the objectives established in the project. Refer also to those objectives that have not been accomplished, justifying the reasons. Organize your report describing the most significant outcomes for the following aspects:

i. Synergy and collaboration among research lines. Also, please explain how FONDAP funding has transformed the individual and collaborative research within the Center?

For the first time, an article has been published that bears the names of all six Principal Investigators of the CGR. This example can be considered a prelude to the multiple papers that are in preparation with the Center's main research lines and that will feature PIs and AIs as co-authors. This trend reflects the continuing desire to work on collective projects and to merge our individual research lines with those chosen at the CGR. This, of course, has only been possible because of the funding provided by FONDAP and the encouragement that becoming part of the Center's main research entails. As proof that this transformation is effectively occurring, many of the new FONDECYT grants obtained by our investigators are focused on aspects derived from our CGR work, indicating a commitment to these lines of research and not just "side projects". The synergy is also manifest in our desire to build a network of centers and to build upon new projects that are more connected with the needs of our country.

In a table and graph shown below, we provide some metrics as to our performance over time. While the number of papers and the impact of these (total and average) are within the normal historical ranges we have had, we note that we have severely overestimated the number of postdocs and students we could sustainably train in our labs. We explain the reasons for this discrepancy in section IV.1.a.ii.

Indicator	2011-2015	Expected for 5 years	2015	2016	2016 Expected	2017	2017 Expected
Number of ISI papers	155	179	39	45	36	41	37
Total Impact Factor of ISI papers	763.833	800	178.053	268.78	178.20	197.376	183.89
Average Impact factor of ISI papers	4.93	4.0	4.57	5.97	4.95	4.81	4.97
Co-authored publications	28 (18%)	20	9 (23%)	10 (22%)	20%	5 (13%)	22%
Postdocs associated to CGR	70	50	40	27	45	29	45
PhD students associated to CGR	73	73	32	29	n/a	22	n/a
Total number of theses directed	136	n/a	83	58	n/a	45	n/a
Co-directed theses	8	12	6	7	10	5	10
Congress presentations	582	n/a	45	93	n/a	69	n/a
Conferences and courses organized	43	n/a	5	6	n/a	12	n/a
Postodcs with permanent positions obtained	10	n/a	0	2	n/a	2	n/a



ii. Formation of advanced human capital directly related to the Center´s objectives: Also, please explain how FONDAP funding has transformed the training of young researchers?

The number of postdocs the CGR had in the first five years was unusually high (40 or so per year) and in 2016 we saw many of the first brood of trainees end their periods with us. Many of them obtained positions in first rate national or international institutions. In addition, about 10 of them continue to be associated to the CGR but we had to change their official affiliation to "young investigator", because of restriction by CONICYT that postdocs can only have funding as such for three years; this increments in about 10 the number of actual postdocs. Regardless, we have about 30 fellows at the postdoctoral level, a number reasonably in line with our previous history. We have no shortage of trainees and, rather than recruiting, we are faced many times with the need to turn applicants away because of lack of additional funds; the FONDECYT postdoctoral program has become highly competitive and there are less funded postdocs available. Nonetheless, we are still receiving very good students at the CGR and we will increase our efforts to incorporate more young scientists to the CGR during 2018.

Our students and young investigators are exposed to world class science, have the chance to create networks with scientists abroad and in Chile, and go on to institutions of excellence to further develop their careers. Perhaps an indicator that could reflect our success in this area would be to have a list of the institutions and centers where CGR students go to after their time with us; we will consider this for future reports.

iii. Collaborative networks both at the national and international level. Also, please explain how FONDAP funding has transformed collaborative research between the Center's researchers and national and international researchers?

The main international networks the CGR has continued to be engaged in during 2017. Among these are those centered at INRIA-Lyon and INRIA-Rennes with which we had multiple exchanges of students, researchers and postdocs and we published 2 common articles. The interactions with colleagues in Germany at the University of Heidelberg were maintained with students being exchanged in each direction, a collaboration based on a joint training project. During 2017 we have also consolidated the scientific work with the Tara Oceans Expedition project. This year the coordinating meeting took held in Santiago and in that occasion we organized and open meeting to discuss about ocean biodiversity and climate change; Professor Karsenti, leader of the initiative has agreed to come back to future meetings that the CGR will organize.

At the national level our main collaborations are with colleagues in other centers of excellence or universities where we are academics. Importantly, we have established a network of five centers (two new ones have joined the initiative in 2017) that will work coordinately in an effort to sequence the genomes of 1000 Chileans and 1000 endemic or economically important species. This project will be launched during 2018 with an outreach activity in schools all over Chile. Further information in Spanish can be found at www.1000genomas.cl (also see below).

iv. Dissemination and exploitation of results

Our investigators have a relationship with different actors in the private sector. Our work Aquainnovo in genetics of salmonids paved the way for participating in the Salmon Genome Consortium and other ventures in the aquaculture field. The main outreach-application of our group is related to the selective breeding and genetic improvement looking for detectable signatures on the genomes of domestic species in the salmon industry together with Aquainnovo. The elucidation of such signatures is fundamental for detecting genomic regions of biological relevance to domestication and then improving management practices in aquaculture. This is also the case with our work on fish pathogens and fish meal development using genetics and animal models. We also have a longstanding relationship with the fruit growers association, the wine industry and the mining industry. Many of these interactions have resulted in results and publications that are of high value to our partners.

Related to our efforts to link science and innovation for the country's needs, we can mention the selection of "FastWine" (by Drs Maass, González and Pulgar) as one of the top 10 Latin American start-ups in the area of agro industry, foods and wine. They have trademarked the name and the company has already started sales.

v. Outreach to society. Please explain the impact of the FONDAP Center in terms of outreach to the general public.

This area is developed in a separate section, which includes all outreach and communication activities (Annex 1).

b. Please explain the impact of the FONDAP Center in terms of contribution to policy makers and other targeted groups.

One of the important decisions we have made as a group of investigators is to explore ways in which to generate a research program that can evolve with the frontiers of science and, at the same time, build on the expertise we have acquired over the years. For this, we are expecting to convince policymakers and agencies interested in R&D that the centers of excellence cannot be abandoned after 10 years of funding. However, we are also aware that extending these programs must be accompanied by a renewal of research goals (to adapt to changing paradigms and frontiers of knowledge) and also of personell and must be merit based. Any type of renewal must involve evaluation of performance and consider the incorporation and mentoring of young scientists that can replace the outgoing generation. We have worked jointly with other centers of excellence to promote these ideas, so far, without much effect in policy.

c. Describe unexpected difficulties encountered and indicate how they were dealt with.

We can describe no significant difficulties or inconveniences during the period reported.

2. RESULTS ACHIEVED PER RESEARCH LINE

Aim 1. A phylogenomic and systems biology approach to identify genes underlying plant survival in marginal soils.

The Atacama Desert is an extremely hostile environment for any living organism. Despite this, a rich plant diversity can be found in this desert. During the last five years, our group has conducted a plant metagenomics study on a transect located in the western slopes of the Andes mountains. The transect spans three environments that can sustain plant life: Prepuna (2.5 - 3.3 thousand meters above the sea level [m.a.s.l.], <30 mm annual rainfall); puna, (3.3 - 4 thousand m.a.s.l., 30-100 mm annual rainfall), and steppe (4 - 4.5 m.a.s.l., 100-200 mm annual rainfall).

We aim to identify Atacama-specific plant adaptations, so, consequently, we have started by a characterizing the exact environmental conditions that surround the Atacama plants under study. We installed two weather stations, at 3060 m.a.s.l. and 4090 m.a.s.l., respectively, and measured weather parameters throughout a year. Precipitations were practically absent throughout most of the year, except for January and February, with 22.6 and 44.4 mm precipitation, respectively, at 3060 m.a.s.l, and 66.4 and 119.2 mm. at 4090 m.a.s.l. Daytime relative humidity was low, with an average of 20.15% at 3060 m.a.s.l and 26.64% at 4090 m.a.s.l. Day-night temperature fluctuations were sharp, with a yearly average of 12.9 ± 2.3 °C at 3060 m.a.s.l., and 13.4 ± 2.7 °C at 4090 m.a.s.l. Daytime radiation levels were extremely high, averaging 620.3 W/m2 at 3060 m.a.s.l and 579.3 W/m2 at 4090 m.a.s.l (global average = 188 W/m2 [1]). We also conducted a general elemental and chemical characterization of the soil throughout the transect. The general soil structure and elemental composition did not show major variations. Its predominant chemical elements were Silicon (Si, ≈60%), and Aluminium (Al, ≈20%), with smaller amounts of Fe, Ca, Mg, K, Ti and other elements. Consistently, its structure was dominated by sand, with smaller amounts of silt and clay. When we analyzed the soluble fraction, we found that Nitrogen, was present in extremely low amounts: The average nitrate concentration was 38.75 mg/kg, equivalent to 91 µM (agricultural concentration=1 - 5 mM [2]), and ammonia averaged 86.82 mg/kg, equivalent to 491µM (agricultural concentration = 7 mM [3]). We also found a pH gradient throughout the transect. It affected the solubility of Fe, Zn, Be, Cu, Mn and P. Soil conductivity values were within agricultural levels at the puna and steppe vegetation belts, however, it was high at the prepuna (up to 4.8 mS/cm, reference value = 2.6 [4]), an indication of high salinity in this area.

We identified sixty-three plant species in the transect, all seed plants. The vast majority of plant species were angiosperms with only one gymnosperm. Our group chose and collected 32 plants that represent the 83% of the total plant coverage throughout the transect and included all plant families found in the transect. Since Atacama soils have low N levels, symbiotic associations with N-fixing bacteria are a likely survival strategy. Consequently, we studied the bacteriome that inhabits the soil immediately adjacent to the chosen plant species. We collected samples of the soil included into the plant´s fine fibrous root system and also samples of bare soil located one meter away from the plant. We extracted soil DNA and used it for taxonomical bacterial identification through microbial 16S ribosomal genes sequencing, and for detection of the nitrogenase subunit nifH, a proxy for bacterial N-fixation. We ribotyped

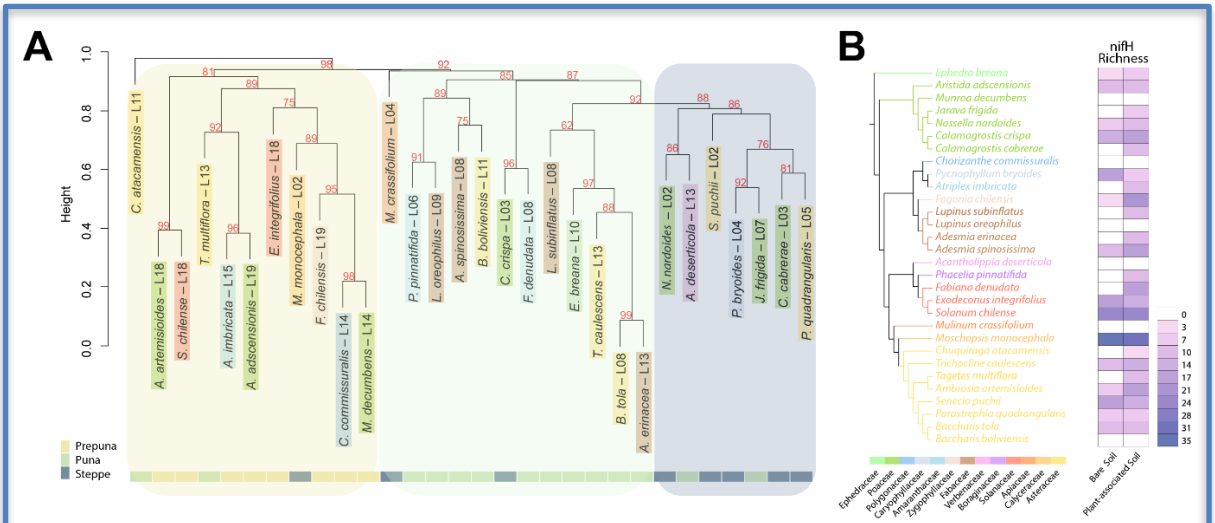


Figure 1. Root-associated Atacama plant bacteriome. A) Cluster dendrogram of the bacterial OTUs associated to the Atacama plants' root system. Each branch contains the plant species name highlighted with a color representing its taxonomic family. Bootstrap values were obtained by 10.000 simulations. The vegetation belt from where the plant soil sample was extracted is shown at the bottom. B) NifH enrichment of plant-associated and bare soil samples. Plant species are shown on the left, with their color representing its taxonomic family. NifH enrichment is shown on the right, with darker colors showing higher nifH enrichment.

3 biological replicates for each plant species and their desert soil controls. We generated 8,936,154 high-quality reads after removal of chimeras and singletons clusters. The most predominant plant-adjacent soil bacterial phyla in Atacama plants were Acidobacteria, followed by Actinobacteria, Armatimonadetes and Bacteroides (43% of all OTUs). The relative abundance of each phyla varied according to the plant species. Bacterial functions were predicted through Functional Annotation of Prokaryotic Taxa (FAPROTAX) [5]. The predominant functions were: General bacterial metabolism (chemoheterotrophy, aerobic chemoheterotrophy, and others), nitrogen metabolism (nitrate reduction, nitrogen respiration, nitrate respiration), nitrogen fixation, and pathogenicity. A clustering analysis of the plant-associated OTUs showed that interspecies bacterial phyla variation follows a geographical pattern (Figure 1A), indicating that plants grown in close proximity share a similar bacteriome. We also surveyed the presence of nifH in both plant-associated and bare soils. As shown in figure 1B, nifH was found in plant-adjacent soil for 80% of the plant species. Plant-adjacent soil of 57% of the tested plant species showed a greater nifH abundance than bare soil. This strongly suggests that N fixation happens in the Atacama plant's rhizosphere, which would confer an adaptational advantage under this environment.

Illumina pair-end sequencing technology was used to characterize the expressed genome of the 32 chosen Atacama plant species. We obtained between 40 to 200 million reads per species. We developed a bioinformatics pipeline for sequence data analysis, assembly and annotation. Technical parameters have already been

described in previous reports. We subsequently built a set of predicted proteins ("proteome") for further analysis. We also identified a set of phylogenetically-related "sister species" and obtained transcriptomic data from public databases. In order to find processes that are present in highly expressed genes of Atacama plants, we performed a gene ontology (GO) biological processes enrichment analysis with the 10% most highly expressed transcripts from each of the 32 Atacama species and 16 sister species. Overall, we found a total of 244 GO terms overrepresented in Atacama plants. We also found that 110 of these terms were overrepresented in Atacama species, but not in their non-Atacama counterparts (One-sided Fisher's exact test $p \leq 0.05$) (figure 2A). A myriad of processes related to osmotic stress, organismal metabolism and transport were overrepresented in the Atacama plants. Interestingly, several GO processes related to N or its metabolites were overrepresented. Other categories were photosynthesis, protein folding and cofactor metabolism. Furthermore, we also tested for processes overrepresented at each individual vegetational belt of the studied transect (One-sided Fisher's exact test $p \leq 0.05$). We only found one process overrepresented on each of the higher belts: Fatty acid biosynthetic process at the puna and response to oxidative stress at the steppe. 26 processes were overrepresented at the prepuna (figure 2B) Processes related to macromolecular complex disassembly, intracellular transport and pyridine-containing compound metabolism are overrepresented in this group. Interestingly, transport, H+-coupled transport and small molecule biosynthesis are processes that contribute to salt and drought tolerance: Enhanced H+ transport has been shown to confer salt and drought tolerance and also increase uptake of the macronutrients K and P. H+ primary transport energizes secondary transport, which contributes to high affinity nutrient uptake and vacuolar compartmentation. Furthermore, synthesis of compatible solutes, such as proline or glycine betaine contributes to water retention in plants that are challenged by arid conditions. These could be contributing mechanisms for plant adaptation to the high salinity and dry conditions imposed by the prepuna.

In order to find evolutionary signatures of adaptation in Atacama plants, our group has analyzed genes undergoing positive selection. The detection of positive selection events is done by calculating the ratio of nonsynonymous mutations per nonsynonymous site to synonymous mutations per nonsynonymous site (dN/dS). Under neutral selection, the nonsynonymous mutations occur at the same rate as synonymous, then $dN/dS=1$. This ratio increases under positive selection. We performed two separate analyses using the SNPGenie software [6]. We started by comparing gene sequences between Atacama species and their sister counterparts. We scored the mean of all pairwise changes at each codon position and compared them against all the possible changes. A Z test was applied for significance. We obtained 1979 positively selected genes. Figure 2C shows a network of positively selected transcription factors.

We also made an analysis consisting on a ten–codon sliding window comparing Atacama plants against their sister species. We obtained a total of 83 genes with at least one window with $dN/dS > 1$. We asked whether any of these genes are involved in any of the overrepresented functions found in the Atacama species (shown in Figure 2A). We looked for the GO terms associated to the positively selected genes and found that seven of them are involved in processes overrepresented in Atacama species (Table 1). We also found an additional subset of nine genes that are involved in

processes overrepresented in Atacama plants that live in the prepuna vegetation belt (Table 2).

In summary, we have done a thorough characterization of the environment surrounding the Atacama plants under study. Our group has gained the first glimpse on the bacterial communities that live in close contact with Atacama plants. Interestingly, the bacterial taxa that we found varies according to geographical positioning and not to plant phylogeny. We also found evidence of N fixation in close proximity to the Atacama plants. Our group has also obtained and analyzed the transcriptome of 32 Atacama plant species. We found that genes involved in metabolism, response to stress, transport and photosynthesis are overrepresented in Atacama species. We also found that transcription factors of the zinc finger, MYB, WRKY, HSF and other families are under positive selection, most likely mediating adaptive changes. Finally, we also found a subset of genes under positive selection that are involved in overrepresented function, most likely mediating important functions in Atacama plants.

Finally, our group is beginning to explore these gene candidates in order to experimentally validate them. We have started to perform experiments with loss-of-function mutants and overexpression in Arabidopsis for the bZIP60 transcription factor, one of the positively selected transcription factors undergoing positive selection in Atacama plants. We are planning to perform validation on additional candidate genes, particularly NTL9, which is not also under positive selection, but also involved in many Atacama-relevant biological processes.

References

1. Rondanelli, R., A. Molina, and M. Falvey, The Atacama Surface Solar Maximum. Bulletin of the American Meteorological Society, 2015. 96(3): p. 405-418.
2. Owen, A.G. and D.L. Jones, Competition for amino acids between wheat roots and rhizosphere microorganisms and the role of amino acids in plant N acquisition. Soil Biology & Biochemistry, 2001. 33(4-5): p. 651-657.
3. Wolt, J.D., Soil solution chemistry: applications to environmental science and agriculture. 1994, New York: John Wiley and Sons.
4. Silva E, H. and J. Rodríguez S, Fertilización de plantaciones frutales. 1 ed. Fertilización de plantaciones frutales. 1995, Santiago, Chile: Facultad de Agronomía, Pontificia Universidad Católica de Chile.
5. Louca, S., L.W. Parfrey, and M. Doebeli, Decoupling function and taxonomy in the global ocean microbiome. Science, 2016. 353(6305): p. 1272-7.
6. Nelson, C.W., L.H. Moncla, and A.L. Hughes, SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. Bioinformatics, 2015. 31(22): p. 3709-3711.
7. Supek, F., et al., REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. Plos One, 2011. 6(7).

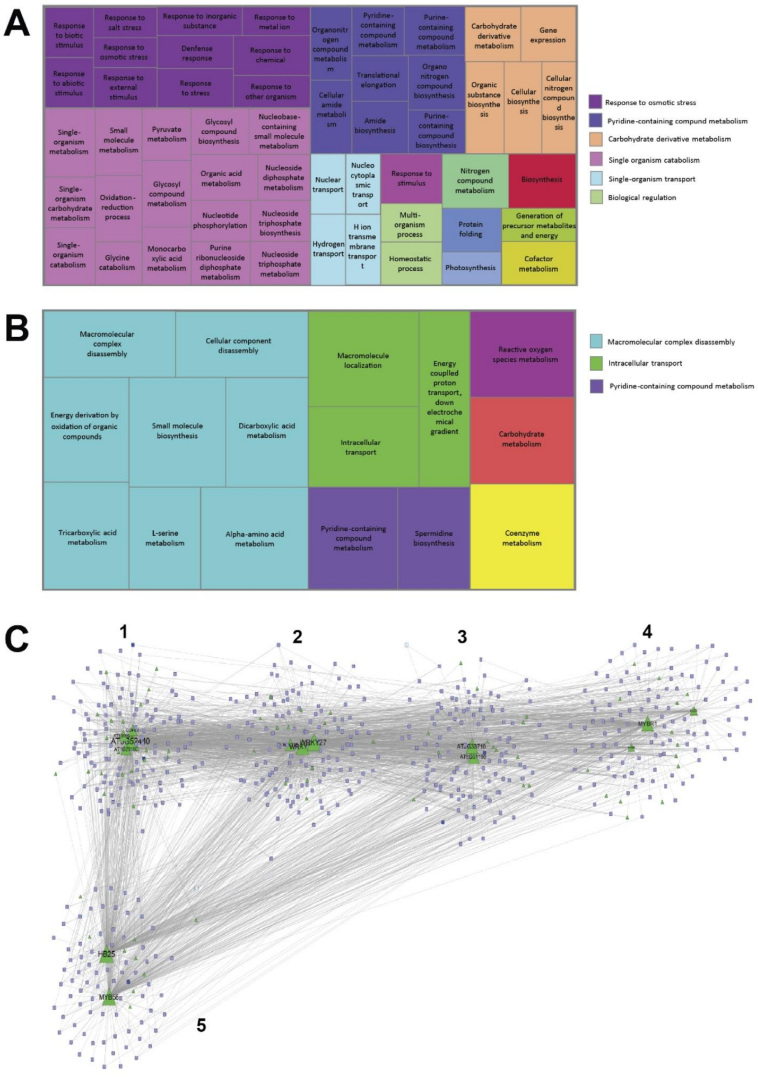


Figure 2. Overrepresented Gene Ontology processes and transcription factors under positive selection in Atacama plants. Processes overrepresented in Atacama plants (A) and the prepuna vegetation belt (B) were analyzed, summarized and grouped by their functional "parent" category with ReviGO [7]. GO terms that belong to the same parent share the same color. "Orphan" categories that do not belong to these functional groups are shown on different colors. A network obtained by DAPseq was constructed with transcription factors undergoing positive selection (C). Transcription factors are shown in green triangles and their targets in blue squares. The size of the triangles is proportional to their connectedness. The transcription factors contained in each cluster are:

Cluster 1: Central nodes: AT3G52440, DOF4, AT1G6957, AT1G29160, AT5G61190. Peripheral nodes: AT1G49850, AGL23, ATH1, AT5G61190, GATA13, AT1G64625, CEST, MYB111, BZIP60, AT3G21330, AT3G60080, AT1G33480, AT4G04580.

- Cluster 2: Central nodes: WRKY29, WRKY27, WRKY55. Peripheral nodes: MYB39, AT3G58720, AT5G02660, AT1G12890, AT5G56980, WRKY35, AT5G22890, AT4G31660, Rap2.6L, BHLH92, AT4G33280.

- Cluster 3: Central nodes: AT2G33710, AT5G51190. Peripheral nodes: AT5G67000, AT4G33940, AT4G38140, NF-YC13, AT2G37580, CRF7, ATS, ERF9, AT3G22100, AT1G19490, AT3G19500, WUE, SBX32, AT4G38140.

- Cluster 4: Central nodes: MYBR1, CUC3, MYB98. Peripheral nodes: AT3G30370, BBX10, AT1G29570, NF-YC10, AT2G03000, AT5G41430, AT3G53310, MYB84, HSFA9, CRF9, UPB1, FD, WRKY10, WRKY43.

- Cluster 5: Central nodes: HB25, MYB58. Peripheral nodes: MYB60, HDG10, ZFP6, LAS, HDG10, WRKY13, AT1G67910, AT4G12670, AT5G15820.

Table 1. Genes undergoing positive selection that are associated to overrepresented functions in Atacama plants. Positive selection was found by comparison of gene sequences of Atacama plants and their sister species. Overrepresented functions correspond to the 10% most expressed transcripts in Atacama plants.

Identifier	Gene	Gene name	Associated Atacama GO terms
AT4G24270	EMB140	Embryo defective 140	GO:0006807 Nitrogen compound metabolic process GO:0010467 Gene expression
AT5G21160	LARP1A	LA related protein 1A	GO:0006807 Nitrogen compound metabolic process GO:0006950 Response to stress GO:0009628 Response to abiotic stimulus GO:0050896 Response to stimulus
AT4G31770	DBR1	Debranching enzyme 1	GO:0006807 Nitrogen compound metabolic process
AT1G75100	JAC1	J-domain protein required for chloroplast accumulation response 1	GO:0009628 Response to abiotic stimulus GO:0050896 Response to stimulus
AT5G59800	MBD7	Methyl-CPG-binding domain 7	GO:0006807 Nitrogen compound metabolic process GO:0009628 Response to abiotic stimulus GO:0044710 GO:0050896 Response to stimulus
AT4G34830	MRL1	Pentatricopeptide repeat (PPR) superfamily protein	GO:0010467 Gene expression
AT4G35580	NTL9	NAC transcription factor-like 9	GO:0006950 Response to stress GO:0006952 Defense response GO:0006970 Response to osmotic stress GO:0009058 Biosynthetic process GO:0009605 Response to external stimulus GO:0009607 Response to biotic stimulus GO:0009617 Response to bacterium GO:0009628 Response to abiotic stimulus GO:0010467 Gene expression GO:0042742 Defense response to bacterium GO:0043207 Response to external biotic stimulus GO:0044249 Cellular biosynthetic process GO:0044271 Cellular nitrogen compound biosynthesis GO:0050896 Response to stimulus GO:0051704 Multi organism process GO:0051707 Response to other organism GO:0098542 Defense response to other organism GO:1901576 Organic substance biosynthetic process

Table 2. Genes undergoing positive selection that are associated with overrepresented functions in prepuna plants. Positive selection was found by comparison of gene sequences of Atacama plants and their sister species. Overrepresented functions correspond to the 10% most expressed transcripts in Atacama plants that live in the prepuna.

Identifier	Gene	Gene name	Associated Atacama GO terms
AT1G18190	GC2	Golgin candidate 2	GO:0046907 Intracellular transport
AT1G25330	CES	CESTA	GO:0044283 Small molecule biosynthetic process
AT2G41730		Unknown protein	GO:0015980 Energy derivation by oxidation of organic compounds
AT3G10572		3-phosphoinositide-dependent protein kinase-1, putative	GO:0046907 Intracellular transport
AT4G13840		HXXXD-type acyl-transferase family protein	GO:0044283 Small molecule biosynthetic process
AT5G43970	TOM22-V	translocase of outer membrane 22-V	GO:0033036 Macromolecule localization GO:0046907 Intracellular transport
AT5G54980	UPF0497	Uncharacterised protein family	GO:0015980 Energy derivation by oxidation of organic compounds
AT5G65410	ATHB25	Homeobox protein 25	GO:0044283 Small molecule biosynthetic process

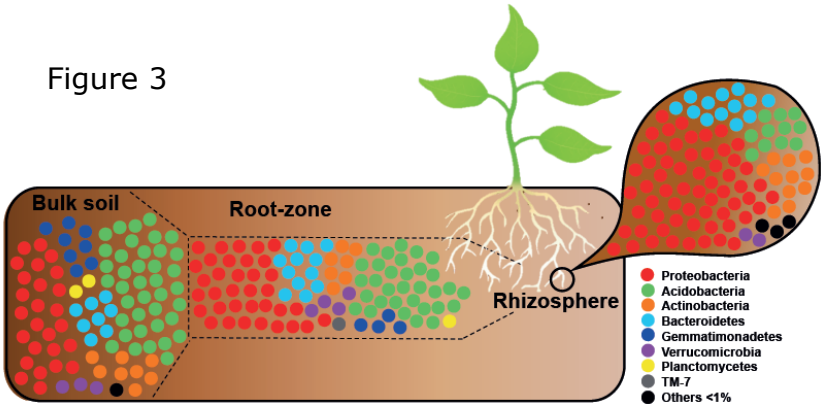
Aim 2. Metagenome of the altiplano soils: plant-microbiome interaction.

The extreme conditions faced by microbial communities inhabiting the Atacama Desert soils provide a unique opportunity to test to what extent microbial community structure is resistant to a strong environmental gradient facing multiple natural stressors, like nutrient-poor soils, extreme solar radiation, large temperature oscillations and elevated salinity, among others. To understand the impact of extreme environmental variables on community structure is a necessary step if scientific studies are to provide answers to global scale issues. During 2017, we examined the microbial community across an altitudinal transect previously named Talabre-Lejía Transect (TLT), where direct anthropogenic interference is minimal to nonexistent. The following topics were developed this year.

Taxonomy and function of the rhizosphere in Andean grassland soils.

The rhizosphere is considered the primary site for soil microbiome differentiation at both taxonomic and functional levels and plant genotype is one of the main drivers in this process. Root-associated bacteria guarantee the survival plants subjected to a great environmental stress. We analyzed the soil microbiome of four native Andean plant species using high-throughput sequencing of the 16S rRNA gene from three soil compartments: the rhizosphere (particles firmly attached to the roots), root-zone soil (loosely attached to the roots), and bulk soil (away from the roots). We compared structure, composition, and function of the microbiome in these three

Figure 3



Taxonomy	Bulk soil*	Root-zone*	Rhizosphere*
Acidobacteria	39.87	32.45	11.09
Actinobacteria	8.24	6.93	7.21
Bacteroidetes	8.39	12.11	13.38
Gemmatimonadetes	8.55	4.00	0.72
Planctomycetes	2.51	1.54	0.74
Proteobacteria	25.15	35.60	61.17
TM7	0.44	1.44	0.80
Verrucomicrobia	3.72	3.69	2.21

*relative abundance (%)

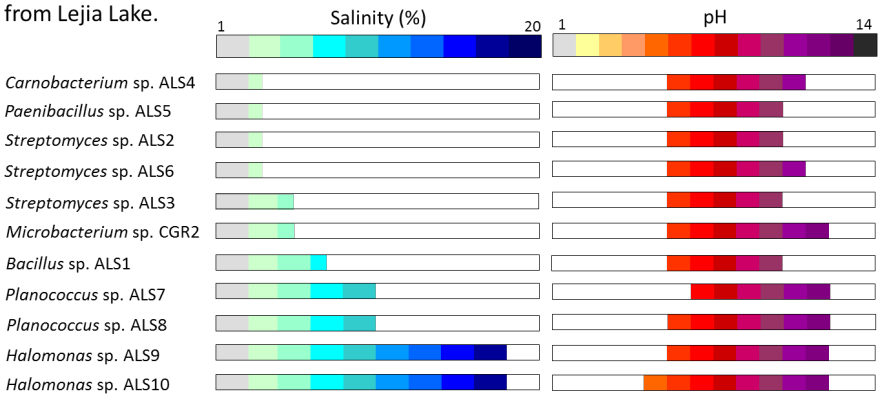
compartments and examined the influence of plant genotype on bacterial communities (Figure 3). Rhizosphere was less diverse than the surrounding soil and taxonomic analyses showed it was enriched in Proteobacteria, while the surrounding soil was enriched in Actinobacteria. Functional analysis did not show clear compartment segregation but showed 23 functions significantly different between rhizosphere and bulk soil. Chemoheterototrophy was the only functional category found significantly more abundant in the rhizosphere. Grasses exhibited a more similar rhizosphere, so we conclude that the influence of plant genotype on the microbiome was stronger at family level rather than at genus level.

Microbiome analysis and isolation of a moderate halophile from Lejía lake soil in Atacama Desert

As a consequence of the severe climate change affecting our entire world, many lakes in the Andes Cordillera are likely to disappear within a few decades. One of these lakes is Lejía lake, which is located at 4,314 meters above sea level (m a.s.l.) in the central Atacama Desert. The objectives developed this year were: 1) to characterize the bacterial community from Lejía lake shore soil (LLS) by using 16S rRNA sequencing and 2) to test a culture-based approach by using a soil extract medium (SEM) to recover soil bacteria. This ecosystem was characterized by an alkaline (pH > 8), saline (4,380 ± 147 mg/Kg of sodium) and nutritionally poor soil (low availability of Fe, Cu, Zn and Mn), dominated by three phyla (Proteobacteria, Bacteroidetes and Firmicutes) that represented over 80% of the OTUs relative abundance. The use of SEM allowed us to recover 7.4 % of the OTUs detected in LLS, all of which belonged to the same three most dominant phyla. Salt supplementation (5 to 20 %) in SEM was used to initiate a selective recovery of salt tolerant cultivable-community members (Figure 4). The genome of one of the most salt tolerant strains was sequenced, which revealed the complete set of genes for polyhydroxybutyrate (PHB) biosynthesis (phaA, phaB and phaC), a metabolic attribute that was verified growing the bacterium at different salt concentrations and using low cost carbon sources. Finally, Halomonas sp. ALS9 transcriptomic response to salt exposure identified specific molecular mechanisms that enable the bacterium to survive under salt pressure.

Figure 4

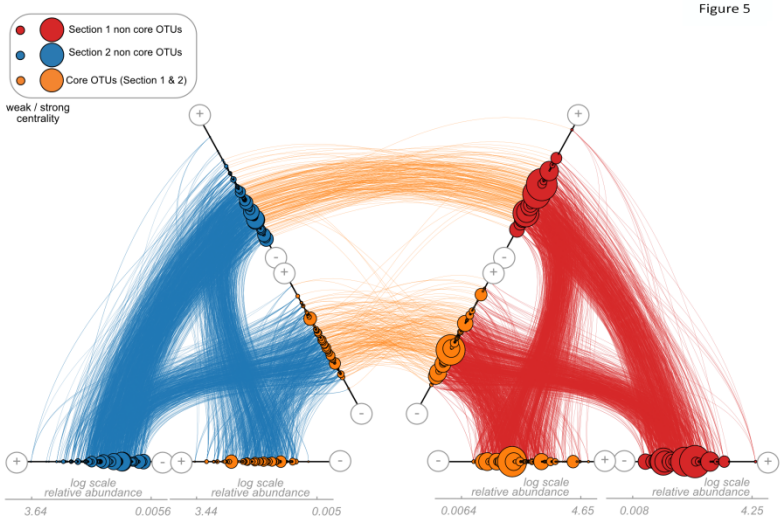
Salinity and pH tolerance tests. These tests were performed to eleven isolates obtained from Lejia Lake.



Structure and co-occurrence patterns in microbial communities under acute environmental stress reveal ecological factors fostering resilience.

Understanding the factors that modulate bacterial community assembly in natural soils is a longstanding challenge in microbial community ecology. In this work, we compared two microbial co-occurrence networks representing bacterial soil communities from two different sections of a pH, temperature and humidity gradient occurring along a western slope of the Andes in the Atacama Desert. In doing so, a topological graph alignment of co-occurrence networks was used to determine the impact of a shift in environmental variables on OTUs taxonomic composition and their relationships (Figure 5). We observed that a fraction of association patterns identified in the co-occurrence networks are persistent despite large environmental variation. This apparent resilience seems to be due to: (1) a proportion of OTUs that persist across the gradient and maintain similar association patterns within the community and (2) bacterial community ecological rearrangements, where an important fraction of the OTUs come to fill the ecological roles of other OTUs in the other network. Actually, potential functional features suggest a fundamental role of persistent OTUs along the soil gradient involving nitrogen fixation. Our results allow identifying factors that induce changes in microbial

Figure 5



assemblage configuration altering specific bacterial soil functions and interactions within the microbial communities in natural environments.

Aim 3. Regulatory landscape plasticity as an evolutionary driver in the genomes of Cyprinidontiform fish.

We have sequenced the genomes of four species of Chilean and two South American Cyprinidontiform fish that inhabit unique and challenging habitats. In the altiplano, members of the *Orestias* genus live in salt pans, high altitude lake remnants that impose different types of stress on the animals including high salt concentration in the water, low oxygen levels and high UV irradiation. Some of these fish live within a few hundred kilometers from the plants and bacteria described in Aims 1 and 2 and, thus, are part of the same biome. We have also examined the genomes of annual fish from the Atlantic coast, animals that display the ability to survive drying out of the ponds they inhabit by laying eggs that can undergo developmental diapause. These fish belong to the genuses *Austrolebias* a *Cynopoecilus* and they have revealed a remarkable genomic evolutionary history which we have discovered through analysis of their sequences. We describe recent progress on both these efforts and we point out that manuscripts describing both projects are in the final stages of preparation.

Orestias ascotanensis.

Orestias ascotanensis (NCBI Taxon ID: 57950), is a teleost killifish (*Cyprinodontiformes*) that is exclusively distributed at the highland systems of the Andes or Altiplano. This fish inhabits the high-altitude lakes (salt and freshwater), specifically in the springs of the Ascotán saltpan in the Chilean Altiplano. Allopatric speciation (geographic speciation) has occurred rapidly in this region due to recent geological changes and sister species have adapted to variable conditions that include high salinity, high altitude, diverse food sources and changing water conditions (e.g., heavy metal content, temperature, etc.). This animal is unique, with characteristics that make it an attractive and ideal model to deepen the understanding of the molecular responses to environmental stimuli that occur in the adaptive process.

Using ALLPATHS-LG 3 we assembled the genome to a size of 696.3 Mb with an average content of GC of 38.63%. Our draft contain 2,394 scaffolds (N50 contig= 43.8 kb; N50 scaffold= 2.67 Mb) with the longest scaffold size of 14Mb. Also, 90% of the genome distributed in 362 scaffolds and the SNP rate was estimated in 1/625. The assembled genome was validated using CEGMA software. We found that 98% of the ultra-conserved core genes for eukaryotes are present in our assembly as complete gene and 100% of the core as partial gene, which correspond to the higher completeness of genome, compared with other teleost fishes. In summary, we identify almost all protein-coding sequences of *Orestias* genome, which implies that our draft represent a high quality and complete genome assembly compared with other sequenced fishes.

We obtained transcriptomic data and we performed an analysis of the ratio of non-synonymous to synonymous between transcript variations identified in RNA sequencing, detecting positive selection pressure in 397 transcripts. Using the

identified transcripts, we performed an ontology analysis which reveals enriched categories related to metal ion binding, DNA repair and response to stress. The list includes the copper-transporting ATPase RAN1, an ortholog of human Menkes and Wilson disease transporters that play a key role in preventing the consequences of copper excess; in addition, we found components for Fe, Zn and Ca uptake: putative ferric-chelate reductase 1-like, zinc transporter ZIP6 (SLC39A6) and the transient receptor potential cation channel subfamily M member 8 (TRPM4). In terms of DNA repair we found the PHR gene, a member of cryptochrome/photolyase family (CPF) that represents an ancient group of widely distributed UV-blue-light sensitive proteins, that play important roles in DNA repair, light perception and circadian clock regulation. Interestingly, the *D. rerio* (zebrafish) ortholog is induced by light and improves UV tolerance. We also found the gene encoding the FANCI protein, which is a component of the nuclear core complex consisting of the Fanconi anemia (FA) protein, that facilitates DNA repair and the PSM1 gene which works together with MLH1 to remove DNA replication errors through a spell-checking mismatch repair mechanism. Interestingly, in yeast exposed to high salt stress, variations of PSM1 contribute to adaptation to changing environments. Finally, we find in the list, the oxygen-dependent choline dehydrogenase (CDHD), fundamental in betaine biosynthesis, a key osmolyte for counteract different environmental conditions (eg. low water, high salinity or extreme temperatures) in different species, through osmoregulation process and used for example as additive in freshwater-seawater transition in salmon farming. Interestingly, betaine also acts as methyl donor and has been studied as regulator of normal methylation in regulation of proto-oncogenes like *c-myc* and also is part of the methionine cycle, fundamental for normal cell growth and proliferation.

In addition to the genome of *O. ascotanensis*, we have sequenced three additional species of this genus: *O. gloriae*, *O. laucaensis* and *O. chungaraensis*. In a second stage of analysis, we will compare the genomes of these species as two of them (*O. ascotanensis* and *O. gloriae*) are salt water adapted while the other two, inhabit fresh water lakes.

***Austrolebias charrua* and *Cynopoecilus melanotaenia*.**

Annual fish are a group of cyprinodontiform teleosts (killifish) that have the unique ability to survive a seasonably variable habitat by specialized adaptations that include a short lifespan and the production of embryos that can undergo developmental diapause, a reversible arrest of embryogenesis that confers tolerance to desiccation for extended periods. It has been suggested that annualism in this group of fish has arisen independently more than once with instances in the Old and New Worlds. In the Neotropics, a major group of annual fish are members of the tribe *Cynolebiini* which originated in the Pampas area in the coast of Southern Brazil, Uruguay and Northern Argentina about 17Mya. Endemic to Uruguay is *Austrolebias charrua* a member of a genus encompassing at least 16 species distributed along the Atlantic coast. These fish inhabit temporary small ponds in open savanna or meadows that dry out during the summer months. Interestingly, the *Austrolebias* genus is characterized by strong genetic flux and variability and a very recent sympatric radiation beginning in the mid to late miocene (11-12Mya). Co-existing with *A charrua* in temporary ponds is *Cynopoecilus melanotenia*, also an annual fish and a member of

the sister tribe *Cynopoecillini*, which diverged from the *Cynolebiini* during the Oligocene (>25Mya), at a time when the common ancestor inhabited the Atlantic coast of Brazil. *C melanotenia* now has a wide distribution and its origin is dated to the early Miocene (15-18Mya) after which it dispersed to its current distribution. Further, *C melanotenia*, as some of its close relatives, has evolved important reproductive differences with other Cyprinidontiforms, including internal fertilization. Further, analysis of DNA content has revealed unusually large genomes in *Austrolebias* (~3Gb), approximately double the size that of *Cynopoecilus* as well as that of other cyprinidontiforms. Importantly, genome amplification has not occurred by multiploidization, as *Austrolebias* are true diploids. These findings suggested that, shortly after the origin of the *Austrolebias* clade, there was a very rapid genome expansion, due to transposable element multiplication, together with a speciation burst.

Whole genome duplication (WGD) has been linked to evolutionary novelty as duplicated genes offer the opportunity for functional diversification. However, genes can also become tandemly duplicated by local chromosomal rearrangements and abnormal recombination events. Further, activation of transposable elements can often lead to gene birth, as retrocopies or genome fragments can be shuttled to new genomic locations. Mobile elements can further contribute to gene diversity by affecting expression of neighboring genes when they transpose to new locations. The selective value of TE insertion near genes has been reported in diverse species. Further, the large proportion of LINE-1, Alu and SVA elements in the human genome, which have expanded over the past 80My of primate evolution, have also been linked to the appearance of specific traits.

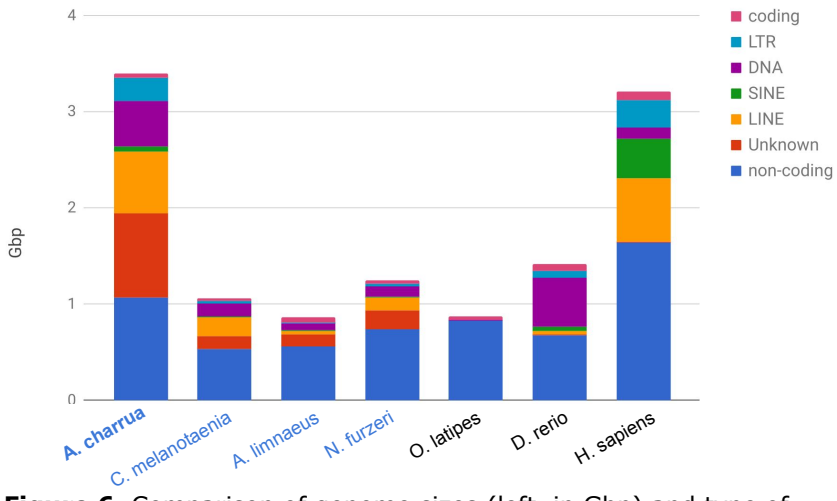


Figure 6. Comparison of genome sizes (left, in Gbp) and type of DNA sequences between fish species and human for comparison. Note that *A charrua* has the largest genome among animals sequenced to date, largely due to repetitive sequences.

We have obtained the complete genome sequence of *Austrolebias charrua* and *Cynopoecilus melanotenia*, two closely related annual cyprinodontiform fish sharing a common environment but displaying markedly different genome structures. While *C melanotenia* has a genome of 1.06 Gb, *A charrua* has a genome size of 3.39 Gb with about 70% of it being repetitive sequence (Figure 6). We identify the classes of mobile elements in their genomes and we show that in *A charrua*, there has been a recent and massive expansion of LINE elements (Figure 7), many of which inserted within or near

genes. Further, we found that *A charrua* has about twice as many genes as *C melanotenia*, suggesting that mobile element expansion impacted gene copy number (Table 3). RNA-seq analysis in both species reveals numerous changes in expression patterns, an indication that gene and TE expansion had functional impact. Finally, we provide data on gene families that show signatures of selection or specific amplification, a gene set that provides clues to understand the life history of this remarkable group of annual fish.

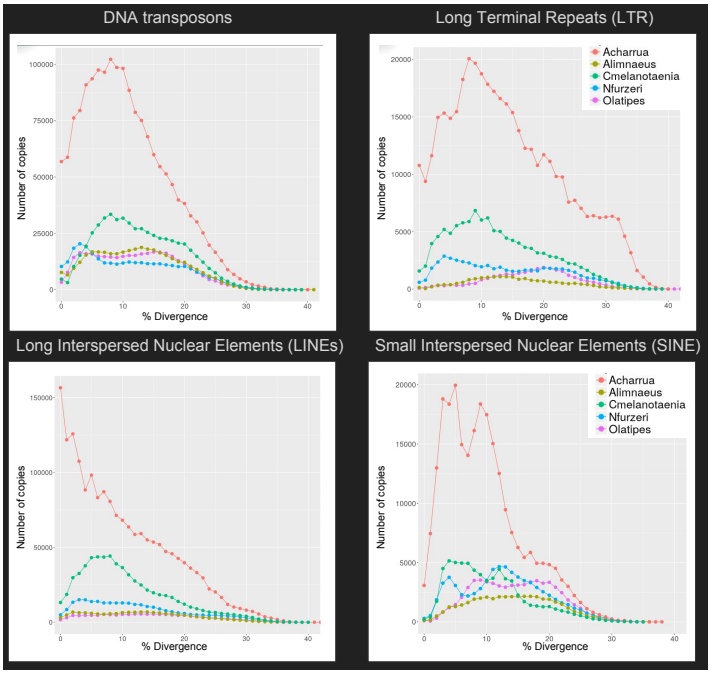


Figure 7. Mobile elements found in the genomes of sequenced fish were aligned and classified according to their percentage of sequence identity. *A charrua* shows not only the largest number of elements but the majority of these show high identity (0-10% divergence) compared to those form other species. This indicates recent expansion of these elements.

Table 3. The genomes of a charrua and C melanotaenia were assembled and annotated revealing that *A charrua* has about twice as many coding genes compared to *C melanotaenia*.

	A. charrua	C. melanotaenia
Number of genes	40,933	20,236
Full length genes	14,942	10,260
Avg. Gene length	1,955.95	1,879.22
Avg. Intergenic length	10,938.54	5,663.17
Avg. Intron length	1,403.32	942.30
Genes assigned to ortholog groups	40,398	20,206
Paralogs rate	2.2	1.58
Exons per gene	6.04	7.25
Genes with complete context (20kb flanking)	4,933	821
Genes with complete context (10kb flanking)	12,989	2,487

Aim 4. Identification of genomic signatures defining metabolic networks that provide unique features to cope with environmental stresses in plants.

Cistanthe longiscapa, a species from the *Montiaceae* family, grows in the Atacama Desert, the driest on earth, during the blooming desert phenomenon, being able to stay alive for a few months under low water availability and high UV irradiation. This species is an outstanding model to study the mechanisms of plants to response to arid conditions. For this reason, our aim in this project is to understand the genetic basis responsible of the *C. longiscapa* ability to survive under extreme arid conditions. Our approach has been based on multidisciplinary strategies in order to characterize *C. longiscapa* considering ecophysiology, genomic, transcriptomic and metabolomic levels.

In this report we are presenting the results associated with the following issues:
1) To characterize ecophysiological traits associated with photosynthesis and aridity in three locations in the Atacama Desert.
2) To improve the genome assembly of *C. longiscapa* using the PacBio platform
3) To identify candidate genes associated with response to abiotic stress

We performed three field trips this year to accomplished with these issues. We took leaf and whole plant samples for DNA and RNA isolation, measured ecophysiological traits from three locations (the same locations that were monitored in 2015) and collected samples to perform an all-day physiological and transcriptomic characterization.

For the first issue, the ecophysiological characterization included measurement of traits associated to CAM photosynthesis like nocturnal malic acid accumulation, chlorophyll and pigment content, leaf water content, leaf mass, leaf carbon isotopic composition, Carbon and Nitrogen composition from plants from the same locations studied during 2015 blooming desert phenomenon. This comparative data will be used to write a manuscript which aim will be to understand the linkage between these two phenomena at the ecophysiological and genetic level, exploring the relation between inbreeding variation, genetic diversity and ecophysiological performance. For genetic diversity and inbreeding estimation, 25 microsatellites will be screening in all the samples collected (10 individuals x 5 locations x 2 years). Currently, samples are being prepared for DNA extraction (removal of cuticule and liophylized). This work is being leaded by a Fondecyt Postdoctoral investigator.

Another interesting trait of *C. longiscapa* is the leaf cuticle, a barrier consisting of lipids and a series of specialized metabolites, that helps leaves to avoid desiccation, provides protection against UV radiation and pathogens attack. The cuticle of *C. longiscapa* is being characterized by means of GC-MS. Alkenes and dehydroabietic acid-like compounds have already been characterized but validation is required. Standards have already been imported and will be used soon.

For the second issue, we have been working to improve version 2 of the genome of *C. longiscapa*. Current plant genomes, published in high impact journals, have a N50 value of above 1.0 x 106 bp. To improve the genome version 2 (N50 = 114.8 x 103 bp), we tested several protocols targeting the extraction of high quality and fragment size DNA. Most of the parameters required for the DNA sequencing by the PACBIO platform, which could have generated the appropriated data for achieving

a high N50, were attained, except for a fragment size above 30 Kb and a 260/230 ratio above 2.0. Thus, we opt to send an individual collected from the 2017 blooming desert for DNA extraction at Polar Genomics (Ithaca, US), for nuclear genome extraction. Over 50 mg of 30-120 kB genomic DNA with a quality above 2.0 and free of RNA contamination was extracted. Half of this amount was sent for PACBIO sequencing at the Icahn Institute and Dept. of Genetics & Genomic Sciences, Icahn School of Medicine at Mount Sinai (NY, US). We expect that the sequencing will conclude during the third week of January 2018. FALCON and FALCON-Unzip software were already installed at the computer facilities of Center of Plant Biotechnology to process the PACBIO data.

For the third point we have focused into generate a *de novo* transcriptome using all the previously generated transcriptome libraries, performed a differential gene expression analysis whithin a sampling design that implied took samples at dawn (8 AM) and dusk (8 PM) in order to evaluate changes in the photosynthetic response within the CAM to C3 spectrum; and to build an all-day transcriptome and physiological analysis in order to characterize completely the photosinthetical cycle. For the *de novo* transcriptome assembling, 34 cDNA libraries from 6 tissues (root, succulent stem, read leaf, green leaf, inflorescence stem and flower) and their technical duplicates (12 in total), seeds (4), and leaves taken at sunset and sunrise from different geological locations (3 plants x 2 moments X 3 sites) sequenced by Illumina with Hiseq 2500 were used. We obtained a transcriptome with 42.133 contigs (N50 1,553 bp), which a 67% were annotated using Blastx and BUSCO analysis identified a 76.5% of ultraconserved genes on the transcriptome. This data will be used to support the gene annotation from the genome and the comparative transcriptome analysis.

Currently, we are finishing the manuscript based on the differential gene expression analysis. This manuscript is based on the ecophysiological and transcriptome data obtained from two sampling locations with different photosynthetic behaviour; we will include both physiological measurements to characterize photosynthesis (acidity, isotopic signature, chlorophyll/carotenoids ratio, leaf mass per area, succulence, C/N ratio, chlorophyll A/B ratio) as well as the transcriptome from dusk (8 PM) and down (8 AM) samples from each location, used for identifying genes with differential expression among the samples to be assessed. In this manuscript, we will show that *C. longiscapa* is able to move along the spectrum between CAM and C3 photosynthesis, and that this behaviour is strongly linked to the geographical location and microclimatic conditions. Currently for this manuscript, qPCR primers are being synthesized.

Finally in this point, 15 whole plants that were collected from a single georeferenced location, at 4 different time-points in one day (15 samples collected at 0 am, 8 am, 12 am, 8 pm, N= 60); will be used to characterize relevant physiological parameters like malic acid accumulation, PEP carboxylase activity, chlorophyll and pigments concentration, carbon isotopic composition; and to prepare pooled cDNA libraries at each time to evaluate genes related to facultative CAM metabolism and processes such as circadian control and response to UV radiation. Samples are being processed in the laboratory and RNA is currently being extracted from these samples.

Aim 5. Gene expression control and regulatory networking.

5.1. Epigenetic mechanisms that control the expression of lineage-specific genes during differentiation.

We have continued our research to assess the principal epigenetic regulators that can "write" and/or "erase" specific histone marks and that can mediate DNA methylation/de-methylation at regulatory regions of the genome, including the Polycomb-Group (PcG), Trithorax/COMPASS (Trx), and Ten-Eleven-Transformation (Tet)-containing complexes. These epigenetic machineries are recruited to target genes where they can either favor silencing or activation of transcription during cell differentiation. In collaboration with other national and foreign groups, we have analyzed the contribution of the these complexes during both neural development and the mesenchymal-osteogenic transition, exploring their specific role during transcriptional control of genes that are critical for establishing and maintaining differentiated phenotypes. We have determined that their activity at target genes occurs concomitantly with the presence of a specific pattern of epigenetic marks at histones as well as with binding of additional epigenetic regulators to promoter sequences that surround the transcriptional start sites (Dudakovic et al. 2017). Additional work between the labs of Drs. Montecino and Palma (former CGR Associate Investigator) that was initiated 4 years ago and published this year (Sepulveda et al. 2017 and Bustos et al. 2017) demonstrated that JARID1B histone demethylase functions as an epigenetic brake for osteogenic-lineage commitment of mesenchymal stem cells derived from the Wharton Jelly of human umbilical cords (WJ-MSCs). This regulatory role is due to the ability of JARID1B to silence transcription of the osteoblast master gene Runx2, as JARID1B demethylates the active transcription-associated mark tri-methyl histone H3 (H3K3me3) and therefore maintains a decreased enrichment of this mark at the Runx2 promoter region. Importantly, these studies indicate that the ability of WJ-MSCs to terminally differentiate to osteoblastic cells also require the full transcriptional activation of the SP7 gene, which codes for a critical transcription factor that controls the expression of several bone phenotypic genes. Interestingly, an international collaboration of Dr. Montecino with Dr. Villagra (George Washington University-USA) supported by the CGR led to a recent publication demonstrating that the SP7 gene transcription during osteogenic lineage commitment of MSCs requires a DNA demethylation step at the proximal SP7 gene promoter mediated by TET-containg complexes that bind to this target region (Sepulveda et al. 2017). Alternatively, in collaboration with Dr. van Zundert at UNAB, Dr. Montecino has assessed regulatory mechanisms that control the function of hippocampal and peripheral-sensory neurons (Bustos et al. 2017a and 2017b; Nuñez-Badinez et al., 2017, respectively). This work led to the discovery that targeted epigenetic editing at the PSD95 gene provides a means to control the expression of this critical protein during hippocampal learning and memory functions. Moreover, by selectively modulating the expression of PSD95, Drs. Montecino and van Zundert were able to diminish the loss of cognitive abilities in Alzheimer mouse models (Bustos et al. 2017b). The work of Dr. Montecino and his team on epigenetic mechanisms within the CRG also made possible the support of research programs of other national investigators, as reflected by two publications in collaboration with these investigators during 2017 (Diaz-Jimenez et al. 2017 and Diaz-Valdivia et al. 2017).

5.2. Biological Network Models

During 2017 we have consolidated (in collaboration with INRIA-Rennes) the development of a pipeline including several new methods for the genome-scale metabolic models reconstruction. Our work proposes an adaptable workspace, AuReMe, for sustainable reconstructions or improvements of genome-scale metabolic models involving personalized pipelines. This ensures that the process is reproducible and documented regardless of the combination of tools used. Additionally, the workspace establishes a way to browse metabolic models and their metadata through the automatic generation of *ad-hoc* local wikis dedicated to monitoring and facilitating the process of reconstruction. Among relevant applications, we illustrate how this workspace allows handling, in an integrated way, the metabolic reconstruction of unexplored organisms such as extremophile bacteria or eukaryote algae. A first example is *Piscirickettsia salmonis*, an intracellular bacterial fish pathogen that causes piscirickettsiosis, a disease with highly adverse impact on the Chilean salmon farming industry. Studies on its metabolism are scarce and only recently a metabolic model for reference strain LF-89 was developed. We present a new genome- scale model for *P. salmonis* LF-89 incorporating specific elements of the fish pathogenic metabolism. Through constraint-based analysis, we determined essential metabolites required for its growth and showed that it can benefit from different carbon sources tested experimentally in new defined media [Cortés et al 2017]. A second model is *Nannochloropsis salina*, a marine microalga, which has become a biotechnological target because of its high capacity to produce polyunsaturated fatty acids and triacylglycerols. It has been used as a source of biofuel, pigments and food supplements, like Omega 3. We present iNS934, the first genome scale metabolic model for *N. salina*. It allowed us to propose 82 different knockout strategies for strain optimization of triacylglycerols [N. Loira et al. 2017]. In the same direction, we have been invited to be part of the third release of the system COBRA (COnstraint-Based Reconstruction and Analysis), which provides a molecular mechanistic framework for integrative analysis of experimental data and quantitative prediction of physicochemically and biochemically feasible phenotypic states. The COBRA Toolbox is a comprehensive software suite of interoperable COBRA methods. It has found widespread applications in biology, biomedicine, and biotechnology because its functions can be flexibly combined to implement tailored COBRA protocols for any biochemical network. Version 3.0 includes new methods for quality controlled reconstruction, modeling, topological analysis, strain and experimental design, network visualisation as well as network integration of chemoinformatic, metabolomic, transcriptomic, proteomic, and thermochemical data.

In the same period we have addressed the reconstruction of networks of microbial interactions with the aim of applying the new methodologies to the data obtained from Atacama CRG-expeditions. For the methodological developments we considered microbial communities in the human gut. One of the principal health care associated infections is the chronic diarrhea generated by the bacterium *Clostridium difficile,* which alters the composition of the gut flora. Nevertheless, little is known about these interactions and properties. To explore the aforementioned interactions, we propose the construction of a neutral space conformed by a set of models that

differ in their interactions, but share the final community states of the gut microbiome under antibiotic perturbation and *C. difficile* infection.

5.3. Complex genome assemblies

During this period we have developed the method FAST-SG, an alignment free method to construct efficient scaffolding graphs to assembly complex genomes using new sequencing technologies. Long read sequencing technologies are the ultimate solution for genome repeats, allowing near reference level reconstructions of large genomes. However, long read *de novo* assembly pipelines are computationally intense and require a considerable amount of coverage, thereby hindering their broad application to the assembly of large genomes. Alternatively, hybrid assembly methods which combine short and long read sequencing technologies can reduce the time and cost required to produce de novo assemblies of large genomes. Our method uses a new ultra-fast alignment-free algorithm specifically designed for constructing a scaffolding graph using light-weight data structures. FAST-SG can construct the graph from either short or long reads. This allows the reuse of efficient algorithms designed for short read data and permits the definition of novel modular hybrid assembly pipelines. Using comprehensive standard datasets and benchmarks, we show how FAST-SG outperforms the state-of-the-art short read aligners when building the scaffolding graph, and can be used to extract linking information from either raw or error-corrected long reads. We also show how a hybrid assembly approach using FAST-SG with shallow long read coverage (5X) and moderate computational resources can produce long-range and accurate reconstructions of the genomes of *Arabidopsis thaliana* (Ler-0) and human (NA12878). FAST-SG opens a door to achieve accurate hybrid long-range reconstructions of large genomes with low effort, high portability and low cost. As a first application we finish the assembly of the Chinook salmon (*Oncorhynchus tshawytscha*), which is an anadromous fish species with considerable ecological, economic and social value, and has been a cultural icon that has sustained native people of western North America for millennia. This species has experienced dramatic long-term declines in abundance due anthropogenic impacts, yet a broad portfolio of phenotypic diversity in natural organisms can buffer against exploitation and increase species persistence in disturbed ecosystems (Schindler et al. 2010). Here we present the first reference genome for Chinook salmon (2.36 GB), with most of the assembly mapped to chromosomes (72.6%) and annotated to enable association mapping of life history variation and phenotypic traits. Whole genome resequencing of populations with distinct life history traits provided evidence that divergent selection was extensive throughout the genome within and among phylogenetic lineages, indicating a broad portfolio of phenotypic diversity exists in this species that is related to local adaptation and life history variation. Association mapping with millions of genome-wide SNPs revealed that a genomic region of major effect on chromosome 28 was associated with both known and cryptic maturation phenotypes during migratory ascent to spawning grounds. Our results demonstrate how genomic resources can enlighten the genetic basis of known phenotypes in exploited species, and assist in uncovering cryptic phenotypes that may be difficult to observe in naturally occurring organisms.

Finally as part of our participation in the consortium for the Atlantic salmon assembly, we were part of the emerging initiative - the 'Functional Annotation of All

Salmonid Genomes' (FAASG), which will leverage the extensive trait diversity that has evolved since a whole genome duplication event in the salmonid ancestor, to develop an integrative understanding of the functional genomic basis of phenotypic variation. The outcomes of FAASG will have diverse applications, ranging from improved understanding of genome evolution, to improving the efficiency and sustainability of aquaculture production, supporting the future of fundamental and applied research in an iconic fish lineage of major societal importance [FAASG Consortium 2017].

5.4. *Novel sequencing platforms to examine the variety of vine genomes.*

Single molecule sequencing has come to revolutionize the genome sequencing field. In particular the MinION platform, uses a completely different sequencing methodology compared to other sequencing technologies, most of them based on sequencing by synthesis. The main advantages of this technology are the removal of PCR amplification and the higher read length compared to other platforms. This last point makes the single molecule sequencing the ideal methodology for sequencing *de novo* genomes. For this reason, at the CGR we have worked on the standardization of the procedures for the construction of MinION sequencing libraries, sequencing and Minion data analysis. To do this, we worked with samples of genomic DNA from *Vitis vinifera* cv. Cabernet Sauvignon using the Ligation Sequencing Kit 1D SQK-LSK108, whereas the sequencing equipment used is the latest version the MinION mk1b with the R9 chemistry.

Three sequencing runs were performed as a proof of concept. The first started with the library construction using 1.2 ug of total genomic DNA without fragmentation. The amount of total DNA at the end of library construction was 236 ng, which was sequenced, and we obtained 34,676 reads (204 Mbp; 0.34X; 15 kbp). For the second run, we started with 2 ug of total DNA performing a fragmentation step. We obtained 246 ng of total DNA at the end of library construction. The sequencing yield was improved, obtaining 55,918 reads and 303 Mbp sequenced, which means a 0.5X of the genome. In the last sequencing run we began the library construction with 4 ug of total DNA which was fragmented. The final amount of total DNA was 522 ng. The sequencing performance was improved compared to the previous sequencing runs. We obtained 108,663 reads and 654 Mbp sequenced, that means a sequencing depth of 1.1X. Once the sequencing runs were performed we evaluated the read length, sequencing quality and sequencing error. The read length and quality were evaluated using the Poretools software. In the first sequencing run the read length had an N50 of 14.4 kbp. For the second sequencing run was 10.7 kbp and in the third sequencing was 11.9 kbp. The average quality of sequencing was 12.4, 10.5 and 11.9 in the first, second and third sequencing run respectively. Finally, the error was evaluated by aligning by Blast against the genome of 'Cabernet Sauvignon', obtaining on average a similarity of 87%, meaning an error close to 13%. Given the high error rate, it is necessary to obtain a greater depth of sequencing in order to correct the error. However, the yield obtained is well below the expected 5 Gbp yield, so it is imperative to continue improving the process of building libraries and sequencing.

V. SUGGESTIONS FROM PREVIOUS EVALUATION

Describe how the suggestions provided by the evaluation panel and the FONDECYT Council in its previous evaluation report were taken into account by the Center.

The evaluation received in mid 2017 corresponding to our 2015-2016 report (we included one and a half years in this report as we had an evaluation at 4.5 years) was overall very positive and encouraging. The peer reviewers had two main points of concern. The first was the need for stimulating further interaction among young investigators at the center. They were satisfied with our progress but proposed to extend our retreat and to have it outside of Santiago. We held the 2017 retreat at the Termas del Corazón for two days and it turned out to be a very valuable instance of discussion. The students themselves made suggestions for improving the event for 2018, for instance, to include a poster session. The second issue had to do with the decrease in the number of postdocs. We have covered this topic above and we expect to continue to improve in this metric.

VI. PRODUCTS GENERATED BY THE PROJECT

In what follows, complete the attached Excel spreadsheets taking into account the following:

REPORT ONLY PUBLISHED MATERIAL INCLUDING THOSE WITH AN OFFICIAL DOI POINTER (e.g., with EARLY ONLINE ACCESS).

EXCEPT FOR BOOKS, ALL BACKUP DOCUMENTS MUST BE PRESENTED IN DIGITAL FORMAT. DO NOT SEND PRINTED COPIES.

ONLY PUBLICATIONS THAT ACKNOWLEDGE THE FONDAP PROGRAM WILL BE CONSIDERED.

1. ISI Publications

- ✓ For each publication, if applicable, the principal author and the corresponding author must be indicated using the following terminology:
  - [1] For principal author (example: Toro[1], J.)
  - [2] For the corresponding author (example: Toro[2], J.)
  - [3] For principal and corresponding author (example: Toro[3], J.)
- ✓ Include a digital copy of each **PUBLISHED** paper.

2. Non ISI Publications

- ✓ For each publication, if applicable, the principal author and the corresponding author must be indicated using the following terminology:

[1] For principal author (example: Toro[1], J.)
[2] For the corresponding author (example: Toro[2], J.)
[3] For principal and corresponding author (example: Toro[3], J.)

- ✓ Include a digital copy of each **PUBLISHED** paper.

3. **Books and book chapters**

- ✓ Include a hard copy of every **PUBLISHED** book.
- ✓ Include a digital copy of the front page of the chapter in the case of a book chapter.

4. **Patents**

- ✓ Include all patents generated by the FONDAP Center.

5. **Congress presentations**

- ✓ Include abstracts of all presentations. Attach a digital copy of the front page of the congress/workshop book.

6. **Organization of Scientific Meetings**

- ✓ List all congresses, courses, conferences, symposia, or workshops organized by the FONDAP Center.
- ✓ Include abstracts of all presentations. Attach a digital copy of the front page of the congress/workshop book.

7. **Collaborative Activities**

- ✓ List the scientific visits of Center members to international institutions
- ✓ List the scientific visits of foreign researchers to the Center in Chile.

8. **Postdoctoral Fellows**

- ✓ List postdoctoral fellows working in the Center during the reported period regardless of their funding sources.
- ✓ Provide current affiliation and positions held by former postdoctoral fellows that left the Center during the reported period

9. **Students**

- ✓ List titles of theses framed in the project completed during the reported period. Attach an abstract and the subject index.

- ✓ List titles of theses in progress, framed in the project, during the reported period. Include digital copies of the corresponding thesis registrations.
- ✓ Provide current affiliation and positions held by former students that graduated during the reported period

10. **Funding Sources**

- ✓ List all funding sources including FONDAP.

VII. OTHER ACCOMPLISHMENTS

Report articles or notes published in the media (provide URL links, if available), awards, prizes, etc.

- During 2017, CRG Principal Investigator Alejandro Maass was elected corresponding member of the Chilean Academy of Sciences.
- Dr. Gutiérrez was distinguished with the John A. Boezi Memorial Alumnus Award for distinguished graduates of the Departament of Biochemistry and Molecular Biology of Michigan State University (USA).
- Dr. Allende was awarded the FEBS Prize for giving the Hermann Niemeyer lecture at the Spanish Society for Biochemistry and Molecular Biology meeting.
- Dr. Allende was part of the official delegation of the University of Chile to MIT and Harvard University in June 2017: http://www.uchile.cl/noticias/133903/rector-vivaldi-continua-viaje-academico-en-harvard-y-mit

VIII. SUGGESTIONS

IX. What recommendations would you make to the FONDAP Program Office to improve the performance of the Center and the review process? Please describe.

X. ANNEXES

1. Outreach

We provide a brief outline of events related to outreach activities, press releases, news items, etc. All accompanying materials are provided in electronic form.

1. Dr. Allende gave the prestigious Bertalanffy Lectures (https://www.cos.uni-heidelberg.de/index.php/COS_Bertalanffy_Lecture) at the University of Heidelberg on October 26-27, 2017. This activity is aimed at local high school students who

have the opportunity to discuss the presenter's work during an entire afternoon (posters in annex).
2. At the outreach level, during 2017, Alejandro Maass was invited to participate as speaker in the Antofagasta Puerto de Ideas Festival, to discuss about climate change and biodiversity. Also the same topic was addressed in the annual workshop of Chilean mathematical engineering students .
3. Postdoctoral fellow Mauricio Latorre carried out an activity with school students at the Liceo Eduardo Charme (San Fernando, VI Region), an activity that had over 200 participants. The title of the talk and discussion was "The importance of metals in Chile" where topics such as iron nutrition, pathogenesis and human health were included.
4. Dr. Maass participated in the "Jornada Tecnológica de Magallanes" in December, 2017, in Punta Arenas. he gave a talk about new opportunities for development with the photonic network available through the optic fiber system in the South of Chile. He also gave a talk on the TARA Expedition and the Oceanic System.
5. Dr. Allende spoke at the 6th Scientific Fair in the city of Osorno. His lecture, attended by about 100 people, referred to the 1000 genomes project (poster in annex).\
6. Dr. Gutiérrez held a "conversatorio" (open discussion with the public) about information management; the title was "¿Ciencia para el bien común? Los desafíos de la transparencia y acceso a información en Chile".

Press articles and appearances (in annex):

1. Article in *Portafolio Financiero*
2. Article in newspaper *La Hora*
3. Article in newspaper *La Tercera*
4. Article in newspaper *Las Ultimas Noticias*
5. Article in *Beaucheff Magazine*
6. Article in *Que Pasa* magazine *http://www.quepasa.cl/articulo/ojos-de-la-llave/2017/06/primer-inventario-de-variedad-genetica-de-chile.shtml/*
7. Interview in *Radio Cooperativa: Audio:* https://www.cooperativa.cl/noticias/sociedad/salud/en-que-consiste-la-edicion-genomica/2017-08-16/112058.html *Video: https://www.cooperativa.cl/noticias/sociedad/salud/ciencia-y-etica-en-la-manipulacion-en-la-correccion-de-defectos-geneticos/2017-08-16/112724.html*
8. Article in the University of Chile web site*: http://www.uchile.cl/noticias/133903/rector-vivaldi-continua-viaje-academico-en-harvard-y-mit*

2. **Other Relevant Aspects**

Our web presence continues with the CGR website (www.genomacrg.cl) and accompanying site such as www.1000genomas.cl and www.atacamagenomes.cl