Comisión Nacional de Investigación Científica y Tecnológica – CONICYT

FONDAP Centers of research Program

Annual Progress Report
2018 (Year 8)

Center for Genome Regulation

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
projas@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

Comisión Nacional de Investigación Científica y Tecnológica - CONICYT

I. PRESENTATION

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

PERIOD COVERED: From: Jan 1, 2018 To: Dec 31, 2018

NAME OF THE CENTER		**CODE**
Center for Genome Regulation (CRG)		15 09 00 07
DIRECTOR OF THE CENTER	**E-MAIL**	**SIGNATURE**
Miguel Allende	allende@uchile.cl	
DEPUTY DIRECTOR	**E-MAIL**	**SIGNATURE**
Martín Montecino	mmontecino@unab.cl	
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: 31/01/2019

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	Fernan 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

2018 corresponds to the eighth year of the Center for Genome Regulation (CGR). Considering the ten year funding period, the CGR is at a critical point in its history: while there is uncertainty as to its continuity (no policy for renewal has been communicated), at the same time it has established a clear direction for the next stage of its scientific development. Further, the period of time that has passed is sufficiently long to be able to thoroughly evaluate the significance of the Center's impact and legacy. This report will show that the trend of increasing scientific relevance and focalization of our research interest will position the CGR as a successful example of how associative funding schemes offer a superb return on investment for the country. Further, we have found a scientific niche that was in dire need of development in Chile and the Region. Worldwide, the field of genomics is showing explosive growth and impact at this moment and we are uniquely positioned to exploit the knowledge base that we have generated. Among other initiatives, we have formalized our commitment together with four other centers of excellence to support the Chilean 1000 Genomes Project, which was launched in mid 2018. The initiative, in turn, joined a worldwide effort aimed at sequencing all eukaryote genomes called the Earth Biogenomes Project (EBP) which was also launched in 2018. Adding to the cutting edge research that is on the way, we have made an effort to involve the wider community in the 1000 genomes initiative by focusing on dissemination and publicity of the idea in the media. Our nationwide school competition (see Section VII), that invited students to participate in a sequencing experiment in the classroom, had huge repercussions and showed an extremely high interest in science projects that are relatable to actual challenges that society faces. This was clearly a learning experience for us scientists as well; the time and effort required to put together a successful outreach program paid off in experience and ability to reach decision-makers in unexpected ways.

Our research during the past year has shown a sustained level of productivity and impact compared with our history. One of the salient highlights has been the publication of our effort to sequence the genomes of persons of Huilliche ancestry, a paper that is authored by all Principal Investigators of the CGR and that went through a very long and exhaustive review process. This work was accompanied by a second publication that relates our genome sequence with specific loci that may have an important consequence for the health of the Chilean population. We have also published articles that address the current principal aims of our project, specifically related to our exploration of the genomes of species inhabiting the Atacama Desert (plants and microbiome). Reports on the genomes of the two groups of fish that we are studying have been submitted and we expect them to be published in 2019. Also, a promising bioinformatic strategy for utilizing sequence data obtained from multiple platforms (short vs long reads) to assemble complex genomes was published in 2018. Added to the above are papers describing new data obtained from economically important species and further advances in our basic research in cell and molecular biology. In total, we published 43 papers (average over 8 years is 36) with a total impact of around 203 (average over 8 years of 183 total impact and slightly over 5.0 average impact).

As has been customary for the CGR's research groups, they continue to attract young talent to their labs. This year we had over 80 young scientists in training working on the diverse projects at the Center and many of them were successful in obtaining their own funding, in a very competitive environment. Many of them also had the opportunity to travel abroad, or to attend important international conferences.

Once again, we have hosted relevant international visiting scientists and organized several workshops and symposia, which further contributed to our students ability to network.

The one aspect of the CGR's activities in 2018 that showed explosive growth is our outreach and dissemination to society. Due mainly to our leading role in the 1000 Genomes Project, we had over 30 media appearances, we were featured in several television or video media programs and we impacted hundreds of school students and their teachers through our workshops, participation in fairs or via the school sequencing activity mentioned above. Further, the CGR Director published a book on genomics in Spanish aimed at the lay person, which was widely distributed for free, mostly to secondary school children. Over 600 copies were printed and given away all over the country. This responded to a decision within our Center to strengthen our interaction with the external environment and promote our science outside of the professional community; we hired a specialized person tasked with leading this effort.

We take the opportunity to highlight two important events and activities that have occurred in early 2019. First, we organized a workshop in epigenomics and related fields in January that was attended by 20 invited speakers from abroad and around 50 participants. The workshop, funded in part by EMBO, included an outstanding group of scientists and generated many new interactions between them and CGR researchers. Secondly, we organized an expedition of 2 young CGR scientists (a postdoc and graduate student) to Antartica, where they have carried out soil sampling and metagenomic sequencing *in situ* at several sites in this continent (see photo). With this new initiative, we are beginning to expand our interest in the microbiome of extreme environments in our country other than the Atacama Desert, where we have thus far concentrated our efforts.

The CGR in Antartica, 2019

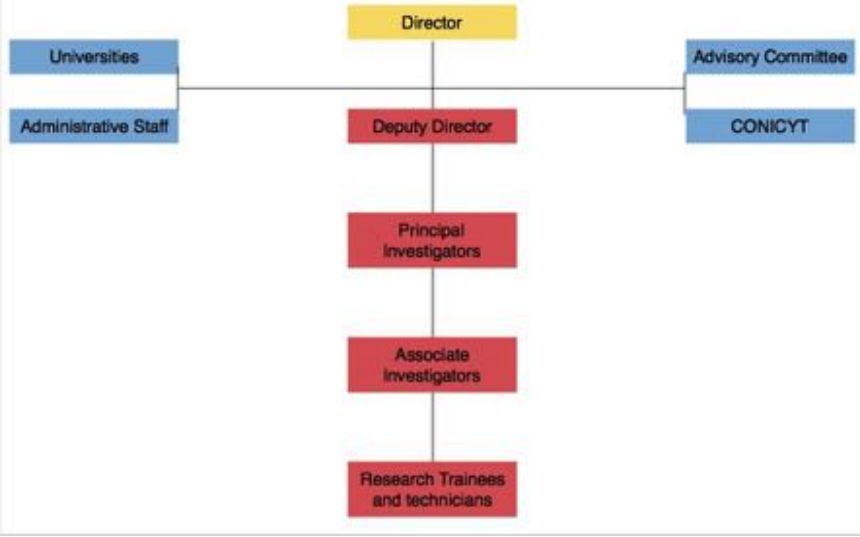
The year 2019 promises to be one of critical decisions and achievements. The most important and anticipated news will be related to the mechanism of funding renewal (or not) for the Center that will be informed to us by the FONDAP Program's authorities. We are fully committed to generating a proposal the guarantees the continuity of our scientific program (which has many ongoing lines of research) as well as to propose new research topics and the incorporation of young talent. The end of our funding period -and hopefully, the beginning of a second stage of development- coincides with the long awaited change in the country's scientific development policy: a Ministry of Science and Technology that will replace CONICYT and other agencies was inaugurated in late 2018. This change brings with it the promise of a new era in which we expect that the centers of excellence are called upon to play a major role. The CGR is ready for such a challenge.

III. ADMINISTRATIVE ASPECTS

1. **Budget execution**: Describe and justify any budgetary modifications (itemized) of the original proposal.
   No significant changes in the budget were needed in 2018.

2. **Accomplishment of institutional commitments:** describe any difficulty(ies) encountered regarding this aspect.
   The institutions that harbor the CGR fully complied with their economic and academic commitments.

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.



Partners:

Chile: ACCDiS, CMM Center for Mathematical Modeling, NLHPC National Laboratory for High Performance Computing Chile, AQUAINNOVO, BIOSIGMA, iBio, Gero

International: MONASH University, ARMI, CNRS, COS Centre for Organismal Studies Heidelberg, UC DAVIS University of California, hhmi Howard Hughes Medical Institute, OpenPlant, EARTH BIOGENOME PROJECT sequencing life for the future of life

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	1	Claudio Latorre	0
	2	Marco Mendez	0
	3	Mauricio Latorre	0
Postdoctoral Fellow	1	Mario Sanchez	44
	2	Ingrid Araya	44
	3	Rodrigo Maldonado	44
	4	Francisca Diaz	44
	5	Soledad Undurraga	44
	6	Grace Armijo	44
	7	Eleodoro Riveras	44
	8	Tomás Moyano	44
	9	Bernardo Pollak	44

	10	Luisa Pereiro	44
	11	Constanza Vásquez	44
	12	Macarena Varas	44
	13	Myra Chávez	44
	14	Jorge Zúñiga	44
	15	Rodrigo Morales Castro	44
	16	Diego Rojas	44
	17	Jonathan Maldonado	44
	18	Dinka Mandakovic	44
	19	Beatriz Fernández-Gomez	44
	20	María Paz Cortés	44
	21	Ángela Cintolesi	44
	22	Paulina Ossa	44
	23	Asier Largo Gossens	44
	24	Dayan Sanhueza	44
Students	1	Sandra Edwards Jorquera	44
	2	Nicolás Cumplido Salas	44
	3	Rodrigo Morales Castro	44
	4	Javiera De la Paz Montt	44
	5	Claudia Molina Pelayo	44
	6	Andrea González Aguilar	44
	7	Karen Fehrmann Cartes	44
	8	Fernanda Lourido Araneda	44
	9	Cristina Muñoz Rehbein	44
	10	Cristina Muñoz Rehbein	44
	11	Miguel Miranda	44
	12	Camilo Valdivieso Guerrero	44
	13	Felipe Gajardo	44
	14	Sebastián Díaz	44
	15	Salomé Muñoz-Sánchez	44

	16	Margarita Carrasco	44
	17	Sebastian Abarzua	44
	18	Esteban Quezada	44
	19	Benjamin Gaete	44
	20	Guillermo Diaz Gamboa	44
	21	Isabel Fredes	44
	22	Tomas Moyano	44
	23	Sebastián Moreno	44
	24	Catalina Ibarra	44
	25	Ariel Cerda	44
	26	Isaak Núñez	44
	27	Tamara Matute	44
	28	María Paz Medina	44
	29	Valentina Zapata	44
	30	Alejandro Fonseca	44
	31	Susan Hitschfelt	44
	32	Jonathan Morales	44
	33	Macarena Muñoz	44
	34	Valentina Zapata	44
	35	Cristopher Hernandez	44
	36	Laura Delgado	44
	37	François Gaspard	44
	38	Isabel Pochet	44
	39	Liliana Lamig	44
	40	Kevin Simpson	44
	41	Isaak Núñez	44
	42	Tamara Matute	44
	43	Gabriel Galvez Jiménez	44
	44	Pablo Orestes	44
	45	Khantati Hauyon	44

	46	Javiera Ortiz Severin	44
	47	Pamela Aravena	44
	48	Alexis Gaete	44
	49	Maria Conztanza	44
	50	Felipe Maza	44
	51	Dante Travisany	44
	52	Martín Rios	44
	53	Diego Gramusset	44
	54	Alonso Espinosa	44
	55	Gerardo Núñez	44
	56	Lissette Ulloa	44
	57	Claudio Urra	44
	58	Anibal Riveros	44
	59	Karin Rothkegel	44
	60	Paula Sandoval	44
	61	Alvaro Barahona	44
Technicians	1	Elvis Acevedo	44
	2	Pamela Vargas	44
	3	Víctor Guzmán	44
	4	Natalia Rojas	44
	5	Diego Henríquez	44
	6	Yariksa Sepúlveda	44
Administrative Staff	1	Florencio Espinoza	44
	2	Carolina Oyaneder	44
	3	Juan Silva	44
	4	Karem Tamayo	44
	5	Priscilla Mardones	44

* Dr. Maass was an Associate Investigator for part of 2018

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

While there have been no changes in identity of our researchers, we had to make one administrative change in order to comply with the guidelines that are followed by both the FONDAP and BASAL programs of CONICYT. Dr. Maass, having assumed (on a rotating basis) the directorship of the Center for Mathematical Modeling (CMM; a BASAL Center), was required to temporarily step down as Principal Investigator of the CGR until he relinquishes the former post. He has thus been re-assigned as an Associate Investigator (which is compatible with his position at the CMM). During part of 2018 and of 2019, we therefore are staffed with 5 Principal Investigators and 6 Associate Investigators, instead of the usual 6 and 5, respectively.

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.

We were able to host our Advisory Board during March of 2018 in Santiago. They met for two days with CGR investigators and personnel and were given all of the materials produced by the Center until that date (Annual Reports and Evaluations). As mentioned previously (2017 report), we have a new member in our Board: Dr. Carlos Pedrós-Alió, who replaced Dr. Ben Koop to avoid a potential conflict of interest (the latter was co-author in an article with one of our investigators). The other two members remain: Dr. Alan Bennett and Dr. Enrique Lessa. The Board provided oral feedback on our projects and strategy and issued a report with their findings and opinions. It was overall positive and we provide it as an Annex in the current report. At the CGR we find their comments extremely useful and they have helped to guide our work for the past years. While all three professors are very productive -and obviously busy- colleagues, they have always shown great willingness to participate in the evaluations and they do so with the rigor and depth that we expect form an internationally recognized panel of peers. We express here our appreciation for their generous work.

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?

In the previous report we informed on the publication of the first article authored by all six Principal Investigators (Pastenes et al., 2017; on the frog *Rhinella*) marking an important milestone in the progress of our Center-wide collaborative projects. This year, we have the second article of this type (Vidal et al, 2018; the Huilliche genomes), a paper that has the six PIs plus four Associate Researchers (two from the previous period). A secongd paper describing data derived from the aforementioned article (Bustos et al., 2019) has also been published and it is co-authored by three CGR investigators. We also have published two papers on the Atacama Desert survey which has now completed its eighth consecutive year. These two articles are authored by multiple CGR investigators. Overall, our collaborative work (papers with more than one CGR investigator) has amounted to over 20% of our papers, the highest ever, and surpassing our prediction for this stage in our history. Overall, the metrics we use to measure our productivity show that, as a group, and with this funding, we perform consistently at a much higher level than was attainable as individuals (see chart in next page). Comparing our 2010 baseline (the situation before the CGR) with the average of the first eight years reveals an increase of 23% in the number of papers, but, importantly, a 54% increase in yearly average impact factor. One can conclude that the quantity, to an extent, and the quality, to a larger extent, were highly influenced by this grant.

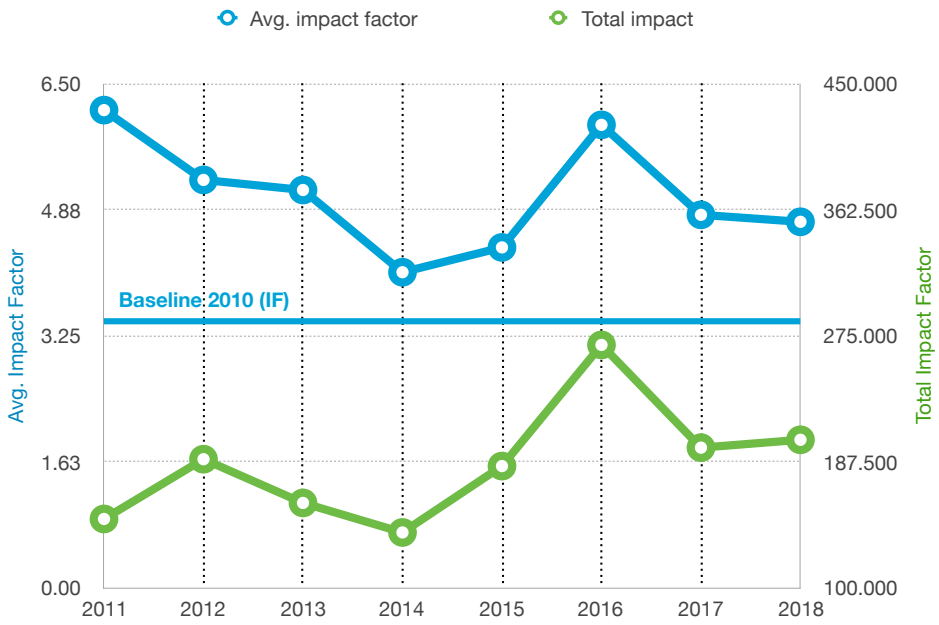
The above not withstanding, it is important to point out that the most collaborative and, hence, impactful research done by the CGR, remains unpublished. Our aims require many years of collection and analysis, and are intensive in experimental and bioinformatic analysis. We predict that the final years and beyond will increase the positive effect of the investment in this group. We remind the reader that our research lines (called "Aims" in our reports and documents) are intrisically synergistic in nature and are not the aims of individual investigators at the CGR; they have been designed from the outset to be the result of contributions from multiple investigators.

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?

In this report, we list the participation of 23 postdocs and 60 students, at different stages in their training. As has been traditional at the CGR, these students have had ample opportunities for interaction in workshops, seminars, congresses and retreats. Many of them have also been able to carry out research stays abroad or to

CGR Productivity

	AVG. IMPACT FACTOR	#PAPERS	TOTAL IMPACT
Baseline 2010	3.28	30	95.000
2011	6.16	24	147.882
2012	5.26	36	189.354
2013	5.13	31	159.007
2014	4.07	34	138.512
2015	4.39	42	184.583
2016	5.97	45	268.680
2017	4.81	41	197.376
2018	4.72	43	202.840
Average	5.06	37	186.029



take part in international meetings and courses. Note that, while in other years we may have had higher numbers of trainees, the impending termination of our funding in 2020 has resulted in our unwillingness to commit to medium to long term hirings and recruitment of personnel.

An interesting fact that emerges from the list of current trainees is that, among our students, 50% are female; in the case of postdocs, the proportion of women exceeds 50%. This had not ocurred before and we are hopeful that a trend to a more balanced distribution of gender is occurring in the field.

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?

One of the main developments in this respect is our decision to become the leading center in the Chilean 1000 Genomes initiative. At the national level, it has allowed us to generate a powerful network of centers (a "supercenter" perhaps) with four other institutions: the Center for Mathematical Modeling (CMM), the Advanced Center for Chronic Diseases (ACCDiS), the Center for Geroscience and Metabolism (GERO) and the Millennium Institute for Synthetic and Systems Biology (iBIO). Further, we have generated a network of associated partners all over the country with which we expect to build a working relationship and to promote genomics as one of the coming revolutions in science. At the international level, it has allowed us to become part of the Earth Biogenome Project, EBP, which aims to carry out a worldwide effort to sequence all eukaryotic life on earth. The CGR signed an M.O.U. with EBP and is now part of this planetary effort in conjunction with numerous leading institutions.

In addition to the above, we joined other international networks that were centered at INRIA-Lyon and INRIA-Rennes with which we had multiple exchanges of students, researchers and postdocs. During 2018 we became part of CNRS Federation Go-SEE. This Research Federation (Ecology and Evolution of Global Ocean Systems) was launched looking for the modeling of the global plankton system. There are eleven international research institutions that seek to understand the fundamental principles of evolution and ecology by modeling life in the oceans and making sense of it through the use of different probabilistic techniques, dynamic systems and data science. The initiative is led by the Tara Foundation, under the supervision of the CNRS, CEA, IRD, five French universities and EMBL.

Interaction with industry: At the national level our main collaborations are with the group of Aquainnovo in genetics of salmonids and with the group of Eduardo Agosín at P. Catholic University in cell factories.

iv. Dissemination and exploitation of results

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

For points iv. and v., please refer to section VII of this report.

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

We have made an effort this year to interact as much as possible with policymakers and the political community in order to make them aware of the potential contribution of our field for the country. We have visited mid level staffers and officials in the ministries of Health, Agriculture and Foreign Relations with this aim. We also had a very productive meeting with the Select Committee on Science and Technology of the Chilean Senate (see Section VII).

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

We can name no difficulties worthy of note during 2018.

2. RESULTS ACHIEVED PER RESEARCH LINE

Briefly describe the main results per research line achieved during the period.

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

1.a. Revealing the Hidden Plant Diversity in Atacama Desert Soils

Worldwide environmental changes are accelerating, and many challenges remain regarding the study and classification of the Earth's vast biodiversity. This is key if we are to improve our conservation efforts and increase our understanding of natural ecosystem functions. Estimating plant diversity in extreme arid environments is particularly challenging, because many plants avoid prolonged periods of extreme drought by setting seed or surviving underground. Due to their sessile nature, the substrate on which plants grow can accumulate seeds, pollen or organic debris that can be used as a proxy to identify the flora present in the ecosystem.

In this study, we used and integrated four different strategies to assess plant diversity in the hyperarid Atacama Desert. We analyzed environmental DNA (eDNA) from soil samples obtained along an elevational transect that spans the limits for plant life in the Andean-Atacama ecosystems. We compared our plant eDNA with a traditional survey performed over eight-consecutive years. Finally, we complement our analyses with soil pollen and seed banks. (Figure 1).

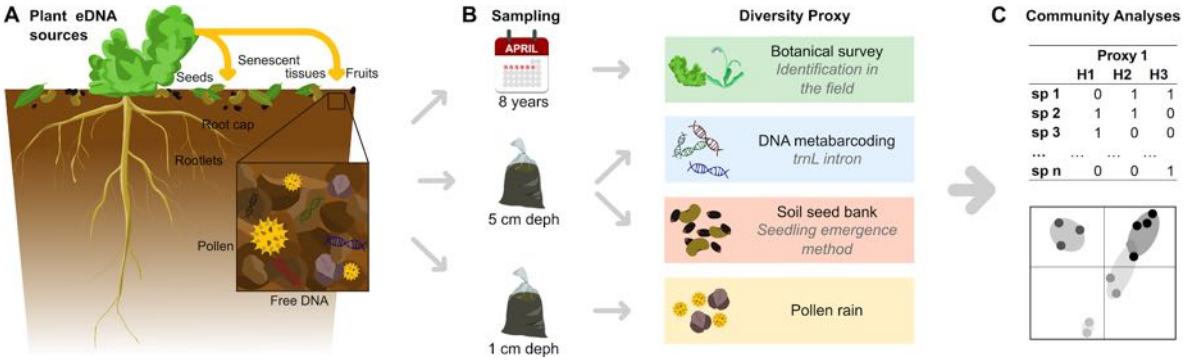


Figure 1. Experimental Strategy. (A) Plant eDNA sources in the soil. Soil contains senescent tissue, seeds, pollen and free DNA. (B) Sampling strategies and diversity proxies. (C) Community analysis.

Soil processing for DNA analysis was made in a PCR cabinet, all surface cleaned with sodium hypochlorite 0.5% and UV light (Illumina, 2013). We extracted total DNA from 10 g of 4 soil sub-samples of each site using FastDNA® 50 mL SPIN Kit for Soil (MP Biomedicals, EEUU). We amplified the P6 loop of the intron in chloroplastid gen *trnL* using *g* and *h* primers (Taberlet et al., 2007). The sequencing was carried out on an Illumina MiSeq (Illumina, San Diego, CA, USA), using the MiSeq Reagent Kit v3 (Illumina) and performing 2 x 201 cycles sequencing, following the manufacturer's instructions. Sequence data was analyzed using OBITools package (Boyer et al., 2016).

Distinct experimental strategies offer different views of biodiversity for a total of 93 taxa identified in this study. We identified 77 plant species over the course of eight years of consecutive surveys through traditional surveys. Observed community composition varies from year to year, with only 22 species observed every year (Figure 2).

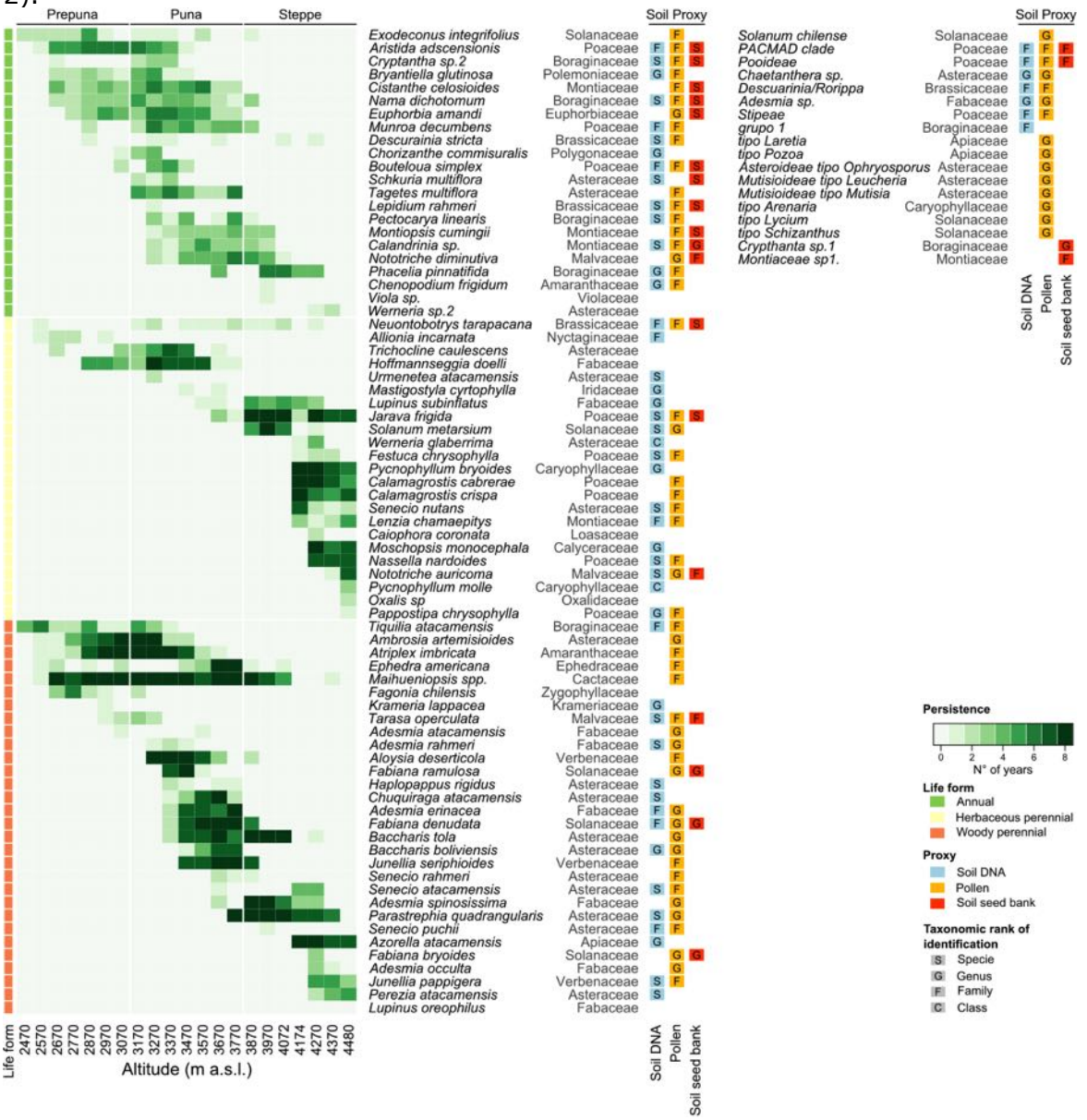


Figure 2. Life form groups the appearance persistence of plants observed in the field, with low persistence in annuals. (A) The left panel shows a heatmap with darker color representing more years of observation of the specie grouped by life form. The right panels of (A) and (B) shows in which soil proxy each taxa appears, and letters within the rectangles represent the taxonomic level in which the specie is identified in every proxy: (S) Specie, (G) Genus, (F) Family and (C) Class. (B) Taxa observed only in soil proxies, but not in the field.

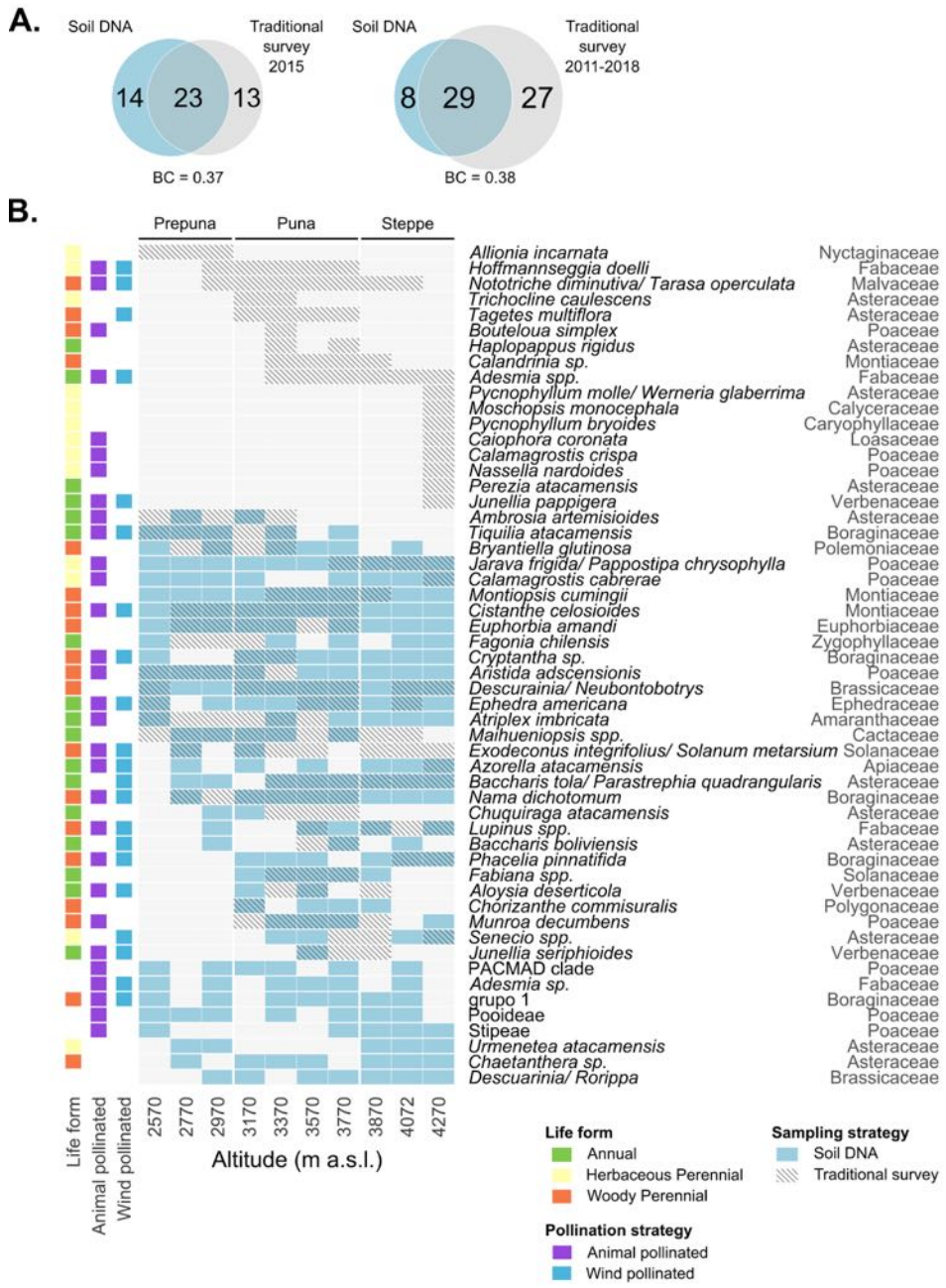


Figure 3. DNA from soil reveal hidden diversity in the transect (TLT) landscape. (A) Euler diagrams comparing at P6 marker taxonomic resolution, between Soil DNA and traditional botanical surveys. The right diagram use 2015 survey, that is the same year of the sampling, meanwhile left diagram compares with the cumulative surveys. (B) Distribution along the TLT of plant taxa in soil eDNA (light blue) compared with botanical surveys plant distribution (striped pattern), plants with no barcode reference sequence were omitted. Life cycle characteristics are shown in the left columns: Life form: Annual (green), Herbaceous perennial (yellow) and Woody perennial (orange). Pollination systems are wind pollinated (blue) and animal pollinated (purple).

Environmental DNA analysis revealed 37 total taxa, seven of which were not observed in our field surveys (Figure 3). Even when diversity described with eDNA metabarcoding is similar the one recorded by the traditional botanical survey (Bray-Curtis dissimilarity index of 0.38, where 0 is two identical sets, and 1 two not overlapping sets), both proxies give complementary information (Figure 3.A). Moreover, soil remains contain eDNA from 14 taxa unobserved the year of the soil collection, of which we have never observed 8 of the taxa in the sites tested (Figure 3.A). Distribution of eDNA in the altitudinal transect is broader than the distribution of the observed above-ground vegetation for most of the taxa observed in these two strategies.

Soil samples from our survey contain a viable seed bank (n = 21 taxa) capable of germinating under laboratory conditions. Soil pollen and eDNA analysis show affinities with vegetation on the landscape but without any obvious relationship to species range or local plot diversity. Most of the observed diversity was retrieved with the complementary proxies (only eight observed species are hidden to the combination of the soil proxies) (Figure 3.A). The taxonomic resolution of the soil proxies varies with the source of the information: pollen mainly identifies taxa at the family level (around a 60% of identified taxa), whereas eDNA and emerged seedlings can be identified to the species level (~42% eDNA and 50% seedlings) (Figure 3). Using eDNA, pollen and the seedlings from the seedbank we observed 18 unobserved taxa (~19% of the complete plant list described here), these elements of hidden diversity have been observed in the region (Marticorena et al., 1998; Rodríguez et al., 2018), so is probable that they could inhabit this transect.

Analyzing beta-diversity of the transect plant communities described with the different proxies (Figure 4) we observed that there is an altitudinal differentiation of the sites described by the long-term plant surveys (higher sites at the left of DCA1, and lower sites at the right in Figure 4). Meanwhile soil eDNA and pollen described community clusters close to the center of the DCA distribution, showing the transect plant community at a landscape scale, which corelates with the broader distribution of the taxa in these proxies. The low richness and highly differentiated sites lead to these communities be apart from the other communities described (Figure 4). The characteristics of each proxy: related with its taxonomic resolution, spatial scale, portion of the diversity retrieved; make each strategy complementary to the other.

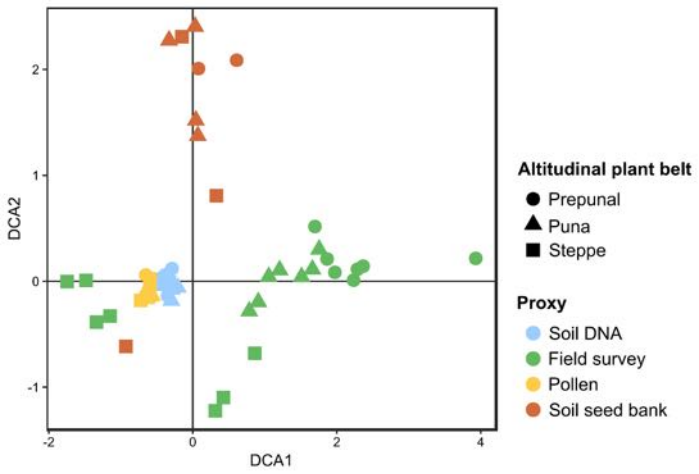


Figure 4. Communities reconstructed from Soil proxies reflect transect (TLT) landscape, and are complementary to each other. Detrended correspondence analysis (DCA). Altitudinal plant belts are represented with different shapes: Prepuna (circle), Puna (triangle), Steppe (square). Meanwhile the proxy from which the community was reconstructed are shown in colors: Soil DNA (blue), Field botanical survey (green), Pollen (yellow) and Soil seed bank (orange). Rare species were downweighted.

Our results reveal how resilient are plant communities in Atacama, with hidden taxa that may have been active in the past or that could emerge under specific environmental conditions in the future. To use eDNA techniques properly is important to assess the advantages and disadvantages of using it in a particular environment. Arid environments seem to be overlooked to molecular analysis of diversity, even when its seasonality makes long term diversity surveys not straightforward, been time and resource consuming to understand its dynamics. Our results show how necessary multiple experimental approaches are for establishing a comprehensive assessment of biodiversity in extreme environments such as the hyperarid Atacama Desert.

References for Aim 1.a.

Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). OBITOOLS: A UNIX-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, *16*(1), 176–182. https://doi.org/10.1111/1755-0998.12428

Illumina. (2013). *16S Metagenomic Sequencing Library Preparation. Part #15044223*. *Illumina.com* (Rev. B). Retrieved from http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/16s/16s-metagenomic-library-prep-guide-15044223-b.pdf

Marticorena, C., Matthei, O., Rodríguez, R., Kalin-Arroyo, M. T., Muñoz, M., Squeo, F., & Arancio, G. (1998). Catálogo de la Flora Vascular de la Segunda Región (Región de Antofagasta), Chile. *Gayana Botanica*, *55*(1), 83.

Rodríguez, R., Marticorena, C., Alarcón, D., Baeza, C., Cavieres, L. A., Finot, V., … Marticorena, A. (2018). Catálogo de las plantas vasculares de Chile. *Gayana Botánica*, *75*(1), 1–430. Retrieved from http://www.efloras.org/madagascar.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., … Willerslev, E. (2007). Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, *35*(3). https://doi.org/10.1093/nar/gkl938

1.b. Multiscale Climate Change Impacts on Plant Diversity in the Atacama Desert

Comprehending ecological dynamics requires not only knowledge of modern communities but also detailed reconstructions of ecosystem history. Understanding the impact of climate change is particularly relevant in arid regions, where ecosystems are often vulnerable to natural or anthropogenic ecosystem changes in water availability. Moreover, plants in such regions display remarkable adaptations to such demanding environmental conditions, and local communities can be surprisingly diverse and unique providing rich genetic resources of global interest. The Atacama is thus an excellent model with which to study biodiversity dynamics and their climatic controls. Using ancient DNA (aDNA) metabarcoding allows biodiversity responses to major climatic change to be explored at different spatial and temporal scales.

Our goal is to understand how different magnitudes of climate change affect plant communities in the Andean Atacama. To integrate spatial and temporal changes in plant biodiversity from annual to millennial time-scales (Figure 5), we compare aDNA data from fossil rodent middens with contemporary plant biodiversity information over the course of eight consecutive years across an altitudinal gradient (from 2,500 to 4,500 m a.s.l) with strong precipitation differences (from 10 to 300 mm/per year). Plant community composition should be affected in proportion to the magnitude of climate change, but the challenge behind this hypothesis is to define the sensitivity of the system. For instance, how has vegetation responded to the last decade of climate change in contrast to much longer timescales with much greater magnitudes of climate change (i.e. across the Last Glacial-interglacial transition) (Nolan *et al.*, 2018).

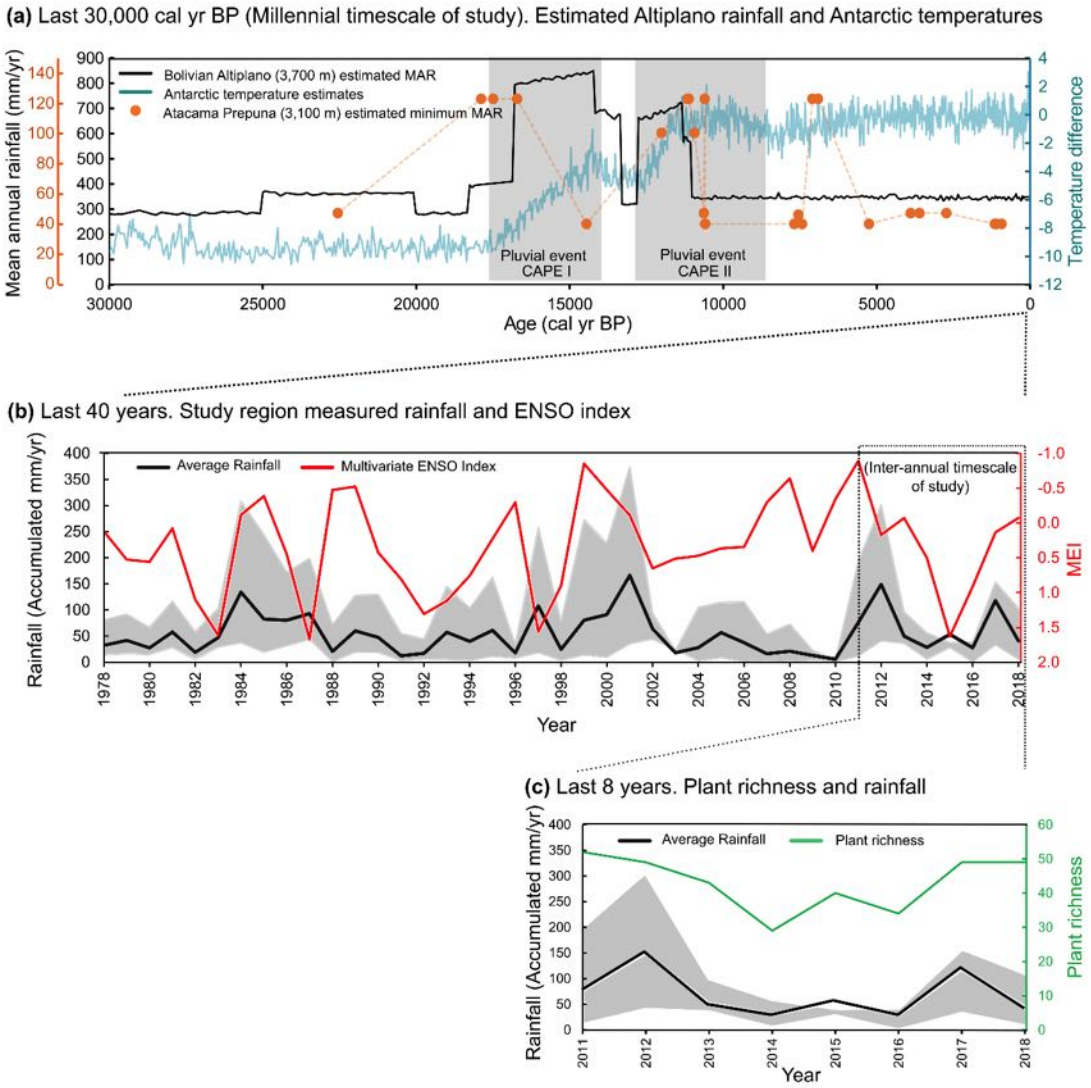


Figure 5: Climate change over the different timescales present in our study. **(a)** Last 30,000 cal yr BP. Black line: Modeled precipitation rates for the Bolivian Altiplano at 3,700 m a.s.l. (m). Rainfall scale (0-900 mm/yr) is shown in black on the left. Orange dots: Minimum Mean annual rainfall (MAR) estimated from fossil vegetation at 3,100 m. Rainfall scale (0-140 mm/yr) is shown in orange on the left. Blue line: Temperature estimated from the Antarctica Epica Dome C (temperature difference from the average of the last 1000 years). Grey boxes: Central Andean Pluvial Event (CAPE) Includes CAPE I (17,500–13,800 cal yr BP) and CAPE II (12,700-8,500 cal yr BP)[(de Porras *et al.*, 2017). **(b)** Last 40 years. Black line: Average total annual rainfall (Data obtained from 20 DGA stations at 22-24° lat S and 2,300-4,400 m, Antofagasta region). Upper grey limit represents average precipitation at 4400-4000 m and lower gray limit represents average precipitation at 2300-2700 m. Red line: Annual Multivariate ENSO Index. Correlation between annual average rainfall and MEI is not significant (r2= 0.04, p>0.05). **(c)** Last 8 years. Black line and gray limits: Same as in B. Green line: Total plant richness observed in the study area after the rainy season (Talabre-Lejía transect, 2500 to 4500 m). Correlation between average rainfall and plant richness is r2= 0.21, p>0.05. Correlation between plant richness and MEI is r2= 0.39, p>0.05.

Present relationships between environmental variables and plant distributions were established with vegetation surveys performed during the first week of April over eight consecutive years (2011-2018) and following every 100 m of elevation (22 sites) across an elevation gradient from approximately 2,500 to 4,500 m a.s.l (Talabre-Lejía transect, TLT). A total of 61 plant species were collected and frozen to construct a local DNA reference database (*Atacama database)* across this gradient.

We extracted aDNA preserved in fossil rodent middens to reconstruct late Quaternary vegetation dynamics in the hyperarid Atacama Desert (Figure 6). By comparing our paleo-informed millennial record with contemporary observations of interannual variations in diversity, we show local plant communities behave differentially at different timescales.

To recover the damaged aDNA and avoid modern DNA contamination, subsampling, extraction and amplification were performed at the specialized aDNA laboratory at Manaaki Whenua - Landcare Research Long Term Ecology Laboratory, Lincoln, New Zealand. To prepare the libraries we followed the protocol "16S Metagenomic Sequencing Library Preparation" (Illumina, 2013), adapted to our marker trnL (c-h). We sequenced using Illumina's next-generation pair-end sequencing technology (MiSeq, Reagent Kit v3, 600 cycles).

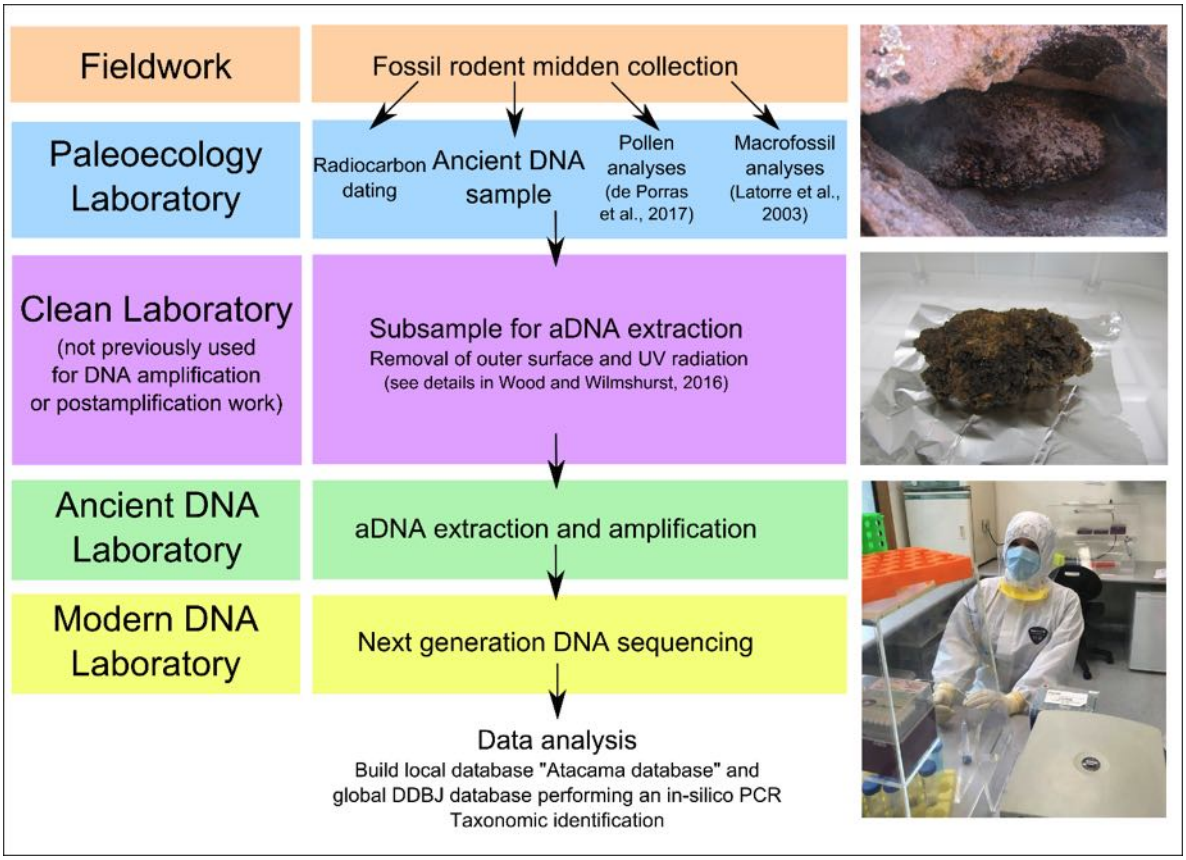


Figure 6: Ancient DNA workflow

We obtained *trnL* sequences or barcodes from 61 species to construct the *Atacama database* (Genbank accession numbers MH115328 - MH115388). The taxonomic resolution of the selected barcode (*trnL c-h)* using only the global *DDBJ database* was around 25% for species, almost 50% for genera and 80% for families. Using our local *Atacama database* the resolution increased to around 80% for species and genera and 100% for families.

To understand diversity changes over longer periods of time than is possible to observe by standard field surveys, we obtained high quality aDNA, amplified and sequenced the *trnL* barcode in fossil samples (dated 200 to 27,590 cal yr BP) and from modern middens using Illumina technology. We obtained a mean of 115,000 reads per midden/sample for 21 middens. 94% of the sequences were taxonomically assigned using both plant databases. 99% of the total assigned reads were classified as Atacama taxa and 1% as contaminants (mainly from vegetation outside the laboratory or food). Using aDNA we identified a total of 73 Atacama plant taxa (25 species, 46 genera and 26 families), including 14 species that were not previously recorded in either pollen or macrofossil records.

As expected, precipitation seems to be the main driver of interannual productivity and plant species richness (Figure 1c). In the inter-annual (years to decades) time frame, only annual herbaceous expand and contract their distributional ranges (emerging from persistent seed banks) in response to precipitation, whereas perennials distribution appears to be extraordinarily resilient (Figure 3b). In contrast, at longer time scales (thousands of years) many perennial species were displaced up to 1,000 m downslope during pluvial events (Figure 3c). At millennial timescales, our results show that large magnitude climate changes (Figure 1) triggered both perennials and annuals to migrate downslope where water limitation and temperatures are extreme today.

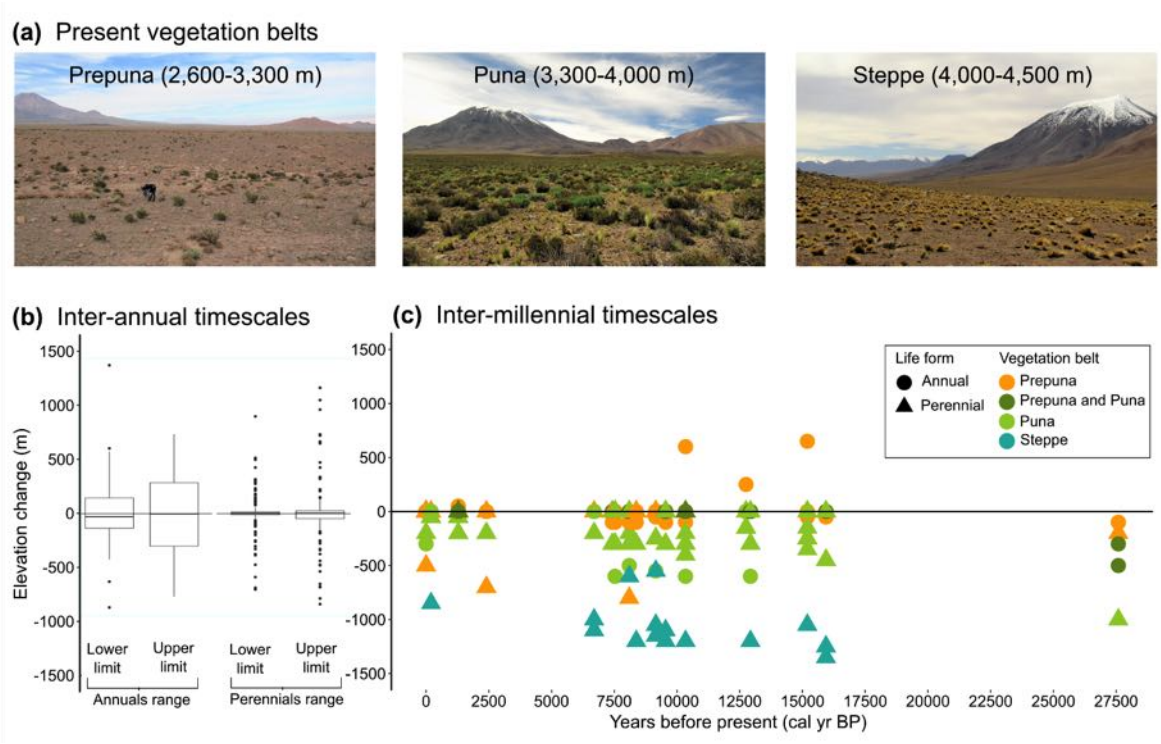


Figure 7 (previous page): Vegetation belts in the Andean Atacama Desert and comparison across different timescales showing major elevational ranges of plant taxa in the study sites. **(a)** Photographs of vegetation belts in the Andean Atacama. **(b)** Boxplots show the interannual variability of range distributions of annual and perennial taxa observed during the eight years of vegetation surveys. Elevation change is shown as annual variations along the upper and lower limits of distributions for each species, compared to the average present-day elevation (averaged over the eight years of the study). Outliers represent species with low coverage (only found in some years). **(c)** Long-term range changes compared to present day ranges (meters below or above their current distribution). Zero here is defined as the maximum distribution range estimated from A. For instance, if a species was found in a 10,000 cal yr BP midden collected at 3,000 m and that species grows today between 4,000-4,200 m, it will appear as a point at -1,000 m (the actual elevational change). As most middens were collected along the lower limit of the vegetation, these limits tend to descend in the past, which gives a good estimate of their past lower limit of distribution.

Although total species richness increased during pluvial events, vegetation belts were not completely displaced downslope. Instead, Prepuna communities were enriched by species moving in from higher elevations (Puna and Steppe) (Figure 4b), generating mixed novel plant communities or "non-analog communities" (e.g.(Williams & Jackson, 2007).

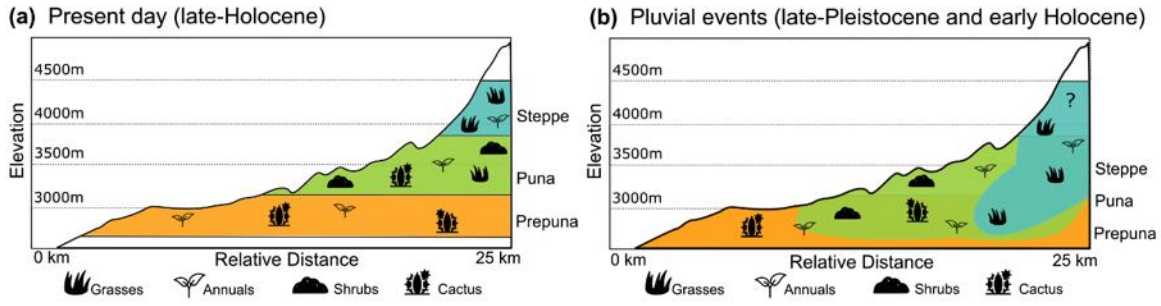


Figure 8: Present day distribution versus major past vegetation change in the Atacama Desert. **(a)** Current vegetation belts at 23°24´ lat. S, Atacama Desert, Chile. As elevation and precipitation increases, temperature decreases. White regions indicate absence of plants. **(b)** Past downslope plant movements during pluvial events provoked a mixing of different vegetation belts or "non-analog communities" from 2500 to 3500 m. The question mark above 4,000 m represents our lack of middens from high altitudes (all midden sites are from 2,414 to 3,380 m, see Table S2 for details).

Long-term paleoecological studies linked to short-term biodiversity studies are fundamental to understanding vegetation responses to ongoing (anthropogenically driven) climate change. The fact that our aDNA analyses were able to replicate and often expand on the total diversity indicates this technology as an important complement to standard midden analysis.

Our integrated approach and data generated should also contribute to understanding the impacts of different magnitudes of climate change on mountain and desert plant communities; vital information for designing conservation management plans and predictive models of global change (Jackson & Blois, 2015).

References Aim 1.b.

De Porras ME, Maldonado A, De Pol-Holz R, Latorre C, Betancourt JL (2017) Late Quaternary environmental dynamics in the Atacama Desert reconstructed from rodent midden pollen records. Journal of Quaternary Science, **32**, 665-684.

Jackson ST, Blois JL (2015) Community ecology in a changing environment: Perspectives from the Quaternary. Proceedings of the National Academy of Sciences, **112**, 4915-4921.

Nolan C, Overpeck JT, Allen JR *et al.* (2018) Past and future global transformation of terrestrial ecosystems under climate change. Science, **361**, 920-923.

Williams JW, Jackson ST (2007) Novel climates, no–analog communities, and ecological surprises. Frontiers in Ecology and the Environment, **5**, 475-482.

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

2.a. Bacterial communities associated to Chilean altiplanic native plants from the Andean grasslands soils.

The rhizosphere is considered the primary place for soil microbiome differentiation and plays a key role in plant survival, especially for those subjected to environmental stress. Using high-throughput sequencing of the 16S rRNA gene, we analyzed and compared soil bacterial communities associated to four of the most abundant high altitude native plant species of the Chilean Andean grasslands. We examined three soil compartments: the rhizosphere (bacteria firmly attached to the roots), the rhizosphere-surrounding soil (RSS, bacteria loosely attached to the roots) and the bulk soil (plant-free soil). The rhizosphere microbiome was in all cases the

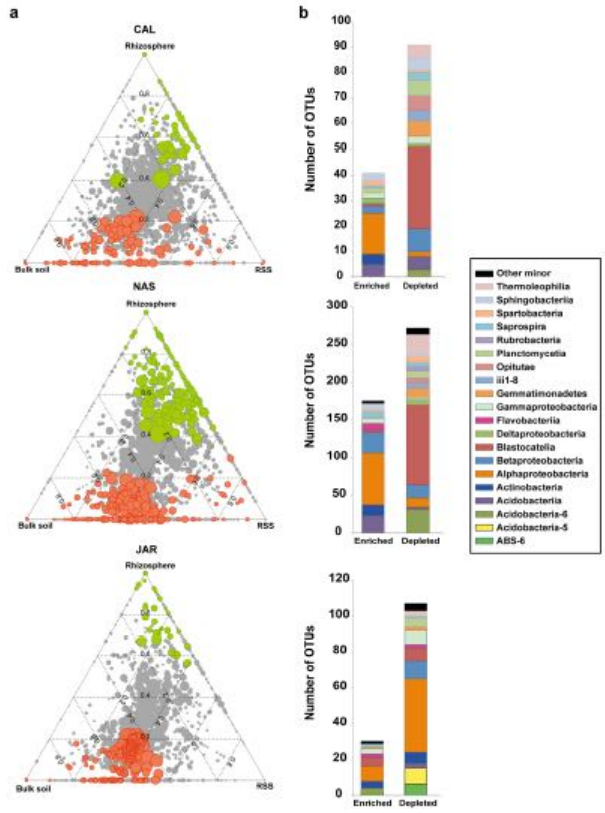


Figure 9. OTU enrichment and depletion in the rhizosphere. (a) Ternary plot representing all OTUs in the data set with relative abundance >0.03% in at least one sample (~ 80% of the total abundance). Dot size represents its mean relative abundance (weighted average) and their position is determined by the contribution of each compartment to its total relative abundance. The dotted grid and numbers inside the triangle indicate 20% increments of contribution from each compartment. Green dots represent OTUs significantly enriched in the rhizosphere. Orange dots represent OTUs significantly depleted in the rhizosphere (FRD; $p<0.05$ in both cases). Gray dots represent OTUs not significantly enriched or depleted. (b) Histograms indicating the taxonomy (class level) of all enriched and depleted OTUs. Only those classes with at least three OTUs in one category (enriched or depleted) are represented. CAL: Calamagrostis crispa, NAS: Nassella nardoides, JAR: Jarava frigida, (Fernandez-Gomez et al., 2019. Scientific Reports. doi: 10.1038/ s41598-018-37776-0.)

least diverse, exposing that the bulk soil was a more complex environment. Taxonomic analysis revealed an abrupt change between the rhizosphere and the rest of the non-rhizospheric soils (Figure 9).

Thus, while rhizobacterial communities were enriched in Proteobacteria (mainly Alphaproteobacteria), Actinobacteria (mostly Blastocatellia) dominated in bulk soils. Finally, we detected certain taxonomic rhizosphere signatures, which could be attributed to a particular genotype. Overall, our results indicate that the thin layer of soil surrounding the roots constitute a distinctive soil environment. This study contributes to expand the knowledge about soil bacterial communities in the Chilean highlands and takes the first step to understand the processes that might lead to the rhizosphere differentiation in that area. Based on our 2018 results we conclude that: 1) native plants might recruit and conserve specific growth-promoting bacteria, allowing them to survive in one of the harshest environments on Earth; 2) that the rhizosphere-surrounding soil (RSS) might be a soil fraction that introduces restrictions and/or promotes the recruitment of a subset of bacteria that colonize the rhizosphere from the surrounding bulk soil, to validate this statement further studies are in progress.

2.b. Soil bacterial communities from the Chilean Andean highlands: taxonomic composition and culturability.

The Atacama Desert is a highly complex, extreme ecosystem which harbors microorganisms remarkable for their biotechnological potential. Here, a soil bacterial prospection was carried out in the high Altiplano region of the Atacama Desert (>3,800 m above sea level; m a.s.l.), where direct anthropogenic interference is minimal. We studied: 1) soil bacterial community composition using high-throughput sequencing of the 16S rRNA gene and 2) bacterial culturability, by using a soil extract medium (SEM) under a factorial design of three factors: temperature (15 and 30 °C), nutrient content (high and low nutrient disposal) and oxygen availability (presence and absence). A

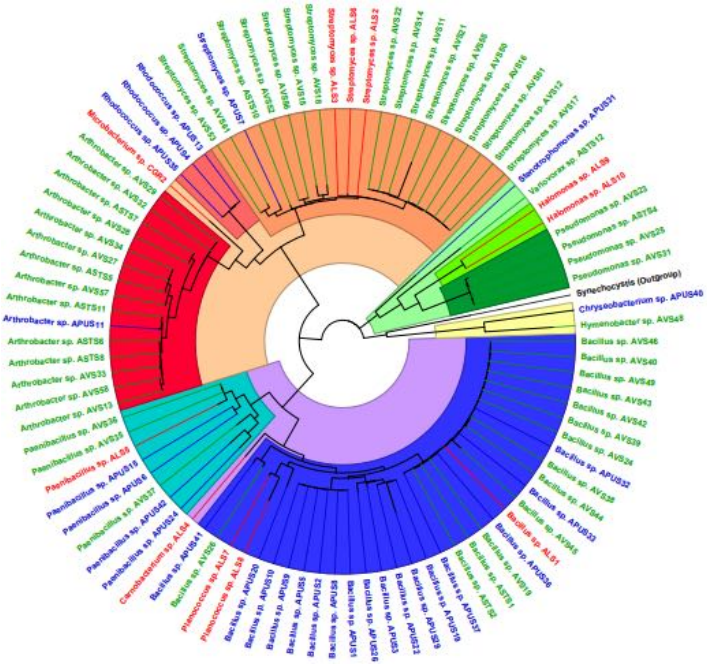


Figure 10: Phylogenetic tree of the bacterial isolates recovered from the highlands of the Atacama Desert. Colors from the inner circle represent the phyla and genera of the bacterial isolates: Proteobacteria (green), Firmicutes (blue specter), Actinobacteria (orange) and Bacteroidetes (yellow); label color represents the isolation site: Lejía Lake (red), TLT1 (green) and TLT8 (blue). Synechocystis was used as outgroup and marked in black. The tree was constructed using the MEGA7 software and edited with the FigTree software (Maza et al., 2019 Front. Bioeng. Biotechnol. doi.org/10.3389/fbioe. 2019.00010).

total of 4,775 OTUs were identified and a total of 101 isolates were selected for 16S rRNA sequencing, 82 of them corresponded to unique or non-redundant sequences (Figure 10). To expand our view of the Altiplano landscape and to obtain a better representation of its microbiome, we complemented our Operational Taxonomic Units (OTUs) and isolate collection with previous data from our group and obtained a merged set of OTUs and isolates that we used to perform our study. Taxonomic comparisons between culturable microbiota and metabarcoding data showed an overrepresentation of the phylum Firmicutes (44% of isolates versus 2% of OTUs) and an underrepresentation of Proteobacteria (8% of isolates versus 36% of OTUs).
Within the Next Generation Sequencing (NGS) results, 33% of the OTUs were unknown up to genus, revealing an important proportion of putative new species in this environment.

Genus	Biomedical	Plant Growth Promotion	Bioremediation
Arthrobacter	-	2	-
Bacillus	3	8	6
Halomonas	-	-	1
Microbacterium	1	1	1
Paenibacillus	2	1	1
Pseudomonas	-	1	1
Rhodococcus	2	1	1
Stenotrophomonas	1	1	1
Streptomyces	9	1	1
Variovorax	-	-	1

Table 1: Most prevalent biotechnological capabilities among the 93 isolates according to the literature. Biochemical characterization and analysis extracted from the literature indicated that an important number of our isolates had biotechnological potential. Also, by comparing our results with similar studies on other deserts, the Altiplano highland was most similar to a cold arid desert. In summary, our study contributes to expand the knowledge of soil bacterial communities in the Atacama Desert and complements the pipeline to isolate selective bacteria that could represent new potential biotechnological resources.

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

3.a. Genome sequence and RNA expression profiles of *Orestias ascotanensis* (Teleostei; Cyprinodontidae) reveal strategies for adaptation to extreme environmental conditions

Orestias ascotanensis (Cyprinodontidae) is a teleost pupfish inhabiting springs in the *Ascotan* saltpan in the Chilean Altiplano (3,700 masl) and is subject to extreme environmental challenges. This species is part of a genus inhabiting both freshwater and salt-lake environments distributed along the central Andes range which shows

multiple allopatric speciation events, most notably in the southern part of its distributional range. We have *de novo* assembled the genome of *O. ascotanensis* at high coverage, representing the first sequenced teleost from a desert environment (see Table 2, for general genome statistics). Comparative analysis of the *O. ascotanensis* genome sequence to those of other previously sequenced teleosts, identifies potential adaptive mechanisms in this species including paralog expansion in families of genes that have been associated with stress resistance to metals ions, salinity and DNA repair after UV exposure, a set that partially overlaps with genes that are under positive selection pressure. We have also found evidence supporting a role for miRNAs expressed in this species, which are predicted to target mRNAs coding for proteins classified in these same categories. Together, our results shed light on the mechanisms operating during adaptive evolution of Andean fishes in response to environmental stress conditions. Furthermore, we propose that these inter-Andean basin lakes represent new natural laboratories suitable for exploring the adaptive strategies developed by their inhabitant species. These results are included in a manuscript that has been submitted for publication (Di Genova et al. see attached copy of the manuscript in the Annex).

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.

Table 2. *O. ascotanensis* genome assembly statistics

Genome Assembly		**N50 (size/number)**	**N90 (size/number)**	**Total length**
	Contigs	43.8 Kb/4,459	11.1 Kb/15,885	670 Mb
	Scaffolds	2.67 Mb/78	383.2 Kb/364	696.3 Mb
Noncoding RNAs		**Copies**	**Length**	
	tRNAs	265	19.6 Kb	
	miRNA	166	12.1 Kb	
TEs		**Total length**	**Percent of genome**	
	Total	142.6 Mb	20.48	
	DNA transposons	39.5 Mb	5.69	
	Retroelements	49.9 Mb	7.19	

Protein-coding genes	**Total Number**	**Annotated**	**Unannotated**
	21,024	19,552	1,072

3.b. Accelerated genome expansion in an annual South American killifish.

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 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 Eastern Brazil about 17Mya. Endemic to Uruguay and Southern Brazil is Austrolebias charrua a member of a genus that encompasses at least 16 species distributed in floodplains along the Southern Atlantic coast (Figure 11A). 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 gene flow and variability and a very recent allopatric 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 their last common ancestor inhabited the coastal plains of Southern Brazil, in the present Atlantic Forest area. C. melanotenia now has a wide distribution (Figure 11A) and its origin is dated to the early Miocene (15-18 Mya) after which it dispersed to its current

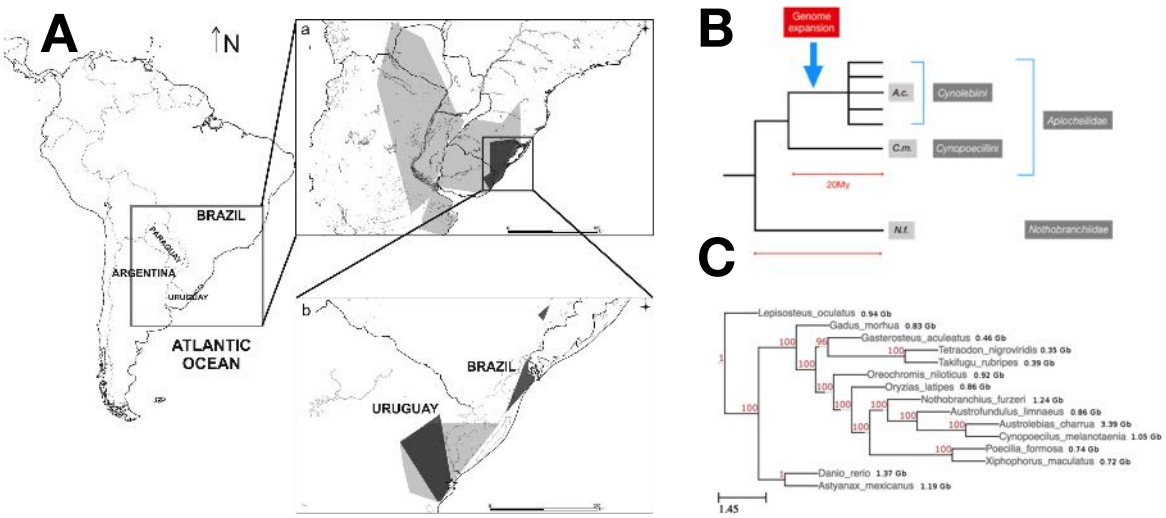


Figure 11. A. Map of South America and two successive insets showing the distribution of C. melanotaenia (gray), A. charrua (black) and the location where both species were collected. B. Schematic tree representing relationships between annual fish of the order *Cyprinodontiformes*. *Nothobranchius furzeri* (N.f.) is a member of the family *Nothobranchiidae* while *Cynolebiini* (includes *Austrolebias charrua*, A.c.) and *Cynopoecillini* (that includes *Cynopoecilus melanotaenia*, C.m.) are tribes in the family *Aplocheilidae*. Arrow indicates branch in which expansion of the genome has occured. C. Phylogenetic tree constructed on the basis of sequence comparisons using 200 conserved genes. Values in red indicate bootsrap values and numbers on the right in black refer to the genome size (in Gb) of each species in the tree.

distribution. Further, *C. melanotenia*, as some of its close relatives, has evolved important reproductive differences with other Cyprinidontiforms, among them, internal fertilization.

Analysis of DNA content has revealed unusually large genomes in *Austrolebias* (~3Gb), approximately double the amount than in *Cynopoecilus* as well as than in other cyprinidontiforms. Importantly, genome amplification has not occurred by multiploidization, as the *Austrolebias* species are true diploids. These findings suggested that, shortly after the origin of the *Austrolebias* clade, there was a very rapid genome expansion likely due to transposable element amplification, an event followed by a speciation burst (Figure 11B).

Whole genome duplication (WGD) has been linked to evolutionary novelty, since second copies of 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.

We have obtained the complete genome sequence of *Austrolebias charrua* and *Cynopoecilus melanotenia*, two closely related South American annual cyprinodontiform fish sharing a common environment and evolutionary history 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. We identify the classes of mobile elements present in both genomes and we show that in *A. charrua*, there has been a recent and massive expansion of LINE elements, many of which inserted within or near genes. Further, we report that *A. charrua* has about twice as many genes as *C. melanotenia*, suggesting that mobile element expansion impacted gene copy number. Tissue-specific RNA-seq analysis in both species reveals changes in gene expression patterns, an indication that gene and TE expansion has 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.

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

The Atacama Desert is one of the driest places on earth, with precipitations only occurring sporadically. When, every few years, thresholds on precipitation abundance and frequency are meet, the phenomenon known as the "blooming desert" is triggered. *Cistanthe longiscapa* (Bernaud) Carolin ex Hershkovitz (Montiaceae) is an endemic annual and one of the most widespread and abundant plants within these blooming desert events in the Atacama. Previous reports based on leaf carbon isotope ratios suggest that Cistanthe longiscapa is in the C3-CAM intermediate spectrum. In order to get further insights, we studied in the field the variation on ecophysiological traits and CAM photosynthesis, reconstructed a *de novo* transcriptome and performed

a RNAseq comparative study from samples taken at the morning and evening to track the changes in CAM/C3 switches under field conditions. Our results indicate that plants from different sites exhibit different levels of CAM photosynthesis. In addition, the transcriptome analysis reveals that plants exhibiting different levels of CAM photosynthesis show different patterns of gene expression (Figure 12). Gene onthology analyses, indicate that categories linked to C3 photosynthesis are more prominent on plants with lower levels of CAM, whereas plants performing higher levels of CAM do not show higher prominence of GO categories related to C3 photosynthesis. Furthermore, plants more engaged in CAM photosynthesis exhibit more expression of genes under the category of "stress". An interesting feature is the expression of a range of genes related to DNA repair and apoptosis on plants from sites exposed to higher stress. The results suggest that *Cistanthe longiscapa* has the ability to respond to different ecophysiological conditions by changing a range of transcriptional networks leading to adaptation.

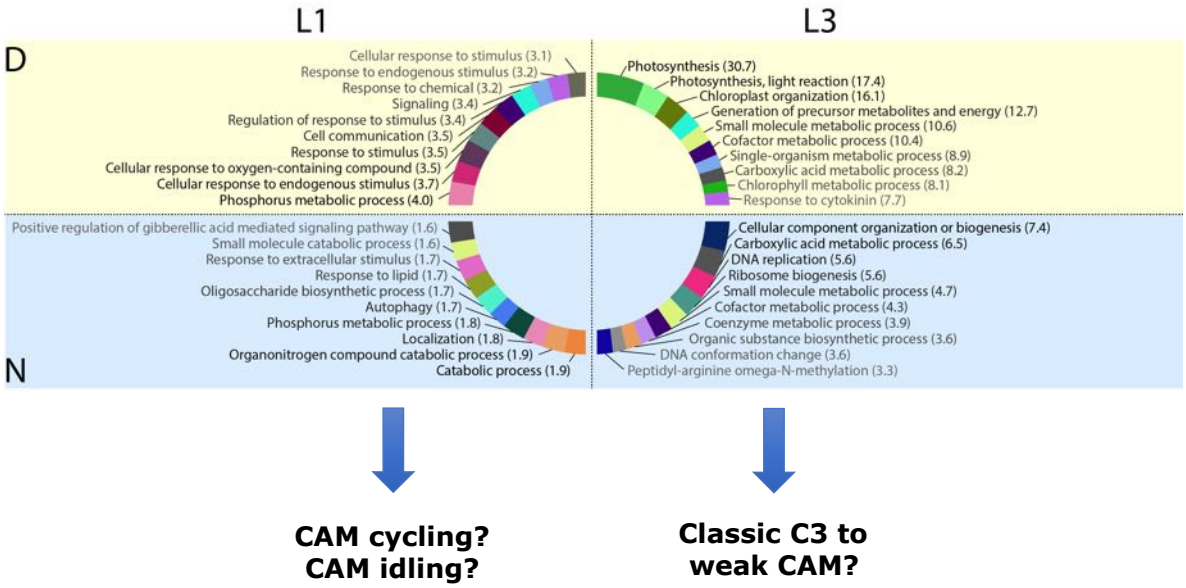


Figure 12. Gene Ontology (GO) enrichment analysis comparing two collection sites (L1, L3) and day (D) vs night (N).

In order to get further insights of the features of this unique desert plant, at the genomic level, we proposed to sequence and perform an assembly *ab initio*. Previous attempts using short read sequences led us to a draft with a scaffold N50 = 114Kb. To improve the metrics of the assembly we change the strategy using long reads (PacBio); thus we obtained an assembly that shows a scaffold N50 = 1.7 Mb which is a substantial increase in the quality of the assembly. Using BUSCO we found around 92% of complete genes, which support the quality of the assembly. Results show a high level of duplicated genes (64%) suggesting that *Cisthanthe* (diploid) suffered a change in ploidy.

This assembly allowed to identify 37,716 protein encoding genes. Currently we are working on the annotation of the genome. In addition, in order to discover how

much genotypic difference occurs among plants collected from different sites in the desert, we have sequenced, using short reads, and we are analyzing the genotypic variability of *Cisthanthe longiscapa*.

PacBio assembly	
Total lenght of sequence (bp)	807,893,784
Total number of sequences	2,254
N50 (contig lenght; bp)	754,944
N80 (contig lenght; bp)	281,748
GC (%)	39
Mean contig size (bp)	358,426.70
Longest contig size (bp)	6,045,800
Contigs > 10 Kb	100
Contigs > 100 Kb	66,02
Contigs > 1 Mb	7,68

PacBio scaffolding	
Total lenght of sequence (bp)	814,341,849
Total number of sequences	1,042
N50 (contig lenght)	1,699,571
N80 (contig lenght)	6,896,198
GC (%)	39
Mean scaffold size (bp)	782,267.21
Longest scaffold size (bp)	9,918,472
Scaffolds > 10 Kb	100
Scaffolds > 100 Kb	79,83
Scaffolds > 1 Mb	24,11

Tables 3 and 4. Genome assembly statistics.

Scaffolding BUSCO	Values	Percentage (%)
Complete BUSCO	1,320	91.6
Complete and single-copy BUSCO	398	27.6
Complete and duplicated BUSCO	922	64.0
Fragmented BUSCO	26	1.8
Missing BUSCO	94	6.6
Total BUSCO groups searched	1,440	100.0

Table 5. Ultra-conserved orthology genes on C. longiscapa genome by BUSCO.
1,440 genes on BUSCO Embryophyta database.
- Complete: complete proteins of BUSCO database found in query
- Fragmented: incomplete proteins of BUSCO database found in query
- Missing: proteins of BUSCO database non found in query

Aim 5. Gene expression control and regulatory networking.

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

We have continued our research to assess relevant epigenetic mechanisms that mediate cell fate decisions in mammalian models. Among these mechanisms those mediated by protein complexes that "write" or "erase" specific epigenetic profiles at regulatory regions of critical genes during these differentiation processes. Most of this work has been carried out in collaboration with research teams at national and foreign institutions, thus not only expanding the limits of our research fields but also providing support, most specially to national groups that do not possess extensive experience in the area of epigenetic control mechanisms. Hence, in collaboration with Dr. Ilona Concha at Universidad Austral de Chile in Valdivia (south of Chile) we determined epigenetic mechanisms mediating the transcriptional responsiveness of Testis Sertoli cells to Wnt/b-catenin signaling. The results, included in a recently published

manuscript (López et al., J. Cell. Biochem. 120: 6753-6762, 2018), established a critical role of this signaling pathway during regulation of the conexin 43 (CX43) gene expression, mostly by mediating the recruitment of histone acetylases to the CX43 gene promoter and up-regulating transcriptional activation of this gene. Also Dr. Allende and Dr. Montecino (both at CRG) collaborated with Dr. Adriana Rojas, at Pontificia Universidad Javeriana of Colombia, to assess the contribution of Mll-COMPASS complexes to generate an epigenetic profile at the proximal regulatory region of the Runx2 gene promoter in pre-osteoblastic cells differentiating to osteocyte-like cells. As Runx2 is the critical master regulator of osteogenic differentiation, this work establishes that binding of Mll-COMPASS complexes to the Runx2 promoter results in the deposit of the epigenetic mark H3K4me3 and transcriptional activation of the gene. The data generated through this collaboration was included in a joint publication (Rojas et al. J. Cell. Physiol. 234: 6244-6253, 2018). Additional mechanisms assessing the role of the transcription factor Runx2 regulating gene expression in bone-related cells were addressed through a collaboration with Dr. Mario Galindo of University of Chile. In this case the role of Runx2 controlling transcription of the ADAM 17 proteinase gene is osteosarcoma cells was established. These results were published as a collaborative work (Araya et al. J. Cell. Biochem. 119: 8204-8219, 2018). In a joint effort with the team led by Dr. Brigitte van Zundert at Universidad Andres Bello, in Chile, it was also addressed whether specific epigenetic mechanisms are associated with variable transcriptional responsiveness of pain-related genes in dorsal ganglia neurons of rats subjected to inflammatory stimuli. The results indicate that groups of these neurons exhibit an epigenetic profile at the promoter region of pain-related genes like P2X3 that facilitate a primary response to inflammation. The results are included in a manuscript published jointly by the teams of Dr. van Zundert and Dr. Montecino (Nuñez-Badinez et al. J. Cell. Biochem. 119: 3922-3935, 2018). Finally, collaborative work between the teams led by Dr. Montecino and Dr. Hodar (both at CRG) allowed, following a combined transcriptomic and bioinformatics approach, the identification of several novel long non-coding RNAs (lncRNAs) that are differentially expressed during osteogenic differentiation. Importantly, one of these new lncRNAs was shown to be critical for the expression of the Runx2 gene, as mentioned before, the master regulator of osteogenesis. Thus, the absence of this lncRNA prevents osteogenesis from occurring. A significant part of these results were included in manuscript published in the period (Nardocci et al. J. Cell. Biochem. 119: 7657-7666, 2018).

5.b. Biological Networks.

During 2018 we have published in PLoS Computational Biology [1] the article consolidating our pipeline and new methods for genome scale metabolic network reconstruction. This is a joint effort with the systems biology group at INRIA-Rennes of the last five years. Our work proposes an adaptable workspace for sustainable reconstructions or improvements of genome-scale metabolic models involving personalized pipelines. A main motivation for developing this pipeline is the metabolic reconstruction of unexplored organisms such as extremophile bacteria found in the CRG Atacama desert expeditions. In particular, through the construction of genome-scale metabolic models (GEMs), in [2] we investigated whether strain-specific features of Microbacterium species were involved in the metabolic ability to tolerate/adapt to local variations within an extreme desert environment. Two Microbacterium sp. CGR1 and CGR2 were studied from two contrasting soil sites in the Atacama Desert: 1) Lascar Volcano and 2) Lejia Lake. Both strains displayed near-identical metabolic

pathways and attributes related to molecular convergence of metabolic capacities, indicating that common environmental features may impose selective pressure on members of this genus. The main difference between the GEMs was in the connectivity of specific metabolites related to pH tolerance and CO2 production, this last one probably used to handle acidic stress through decarboxylation reactions. A HPLC assay highlighted the presence of β-carotene, a molecule widely used in biotechnology applications. This study provides a valuable resource to further investigate global metabolic adaptations of bacterial species to grow in soils with different abiotic factors within a common extreme environment. Finally, concerning metabolic networks we were asked to write a review [3] of recent constrained optimization frameworks for pathway selection, as well as relevant pathway engineering case studies that highlight the importance of rational metabolic designs and reconstructions. Indeed, design and selection of efficient metabolic pathways is critical for the success of metabolic engineering endeavors. Convenient pathways should not only produce the target metabolite in high yields, they also require to be thermodynamically feasible under production conditions, make efficient use of available redox and energy cofactors, and prefer efficient enzymes. To support the design and selection of such pathways, different computational approaches have been proposed for exploring the feasible pathway space under many of the above constraints. We also discuss the challenges and limitations hindering the broad adoption of these methods in metabolic engineering projects.

The methodological study of the reconstruction of the dynamics of a metagenomics sample was undertaken in [4] using as example the more common healthcare associated infection caused by the bacterium *Clostridium difficile* which alters the normal composition of the human gut flora. We propose a method that staring from a Boolean network associated to the dynamics of a microbial community explores the space of Boolean networks sharing a number for dynamical features with the initial one looking for the best network in relation to a given phenotype. In the case of the studied infection we were able to reconstruct a better dynamical network than the one described in literature in the sense that much more experimental observations are captured.

5.c. Complex genome assemblies.

During this period the method FAST-SG, an alignment free method to construct efficient scaffolding graphs to assembly complex genomes using new sequencing technologies was published in the journal Giga Science (Di Genova et al., 2018). This method is going to be the base of our assemblies of complex genomes in the 1000 Genome project developed at the CRG.

5.d. Applications.

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 aquaculture together with the company Aquainnovo. Selective breeding and genetic improvement have left detectable signatures on the genomes of domestic species. The elucidation of such signatures is fundamental for detecting genomic regions of biological relevance to domestication and improving management practices. In aquaculture, domestication was carried out independently in different locations worldwide, which provides opportunities to study the parallel effects of domestication on the genome of individuals that have been selected for similar traits. In [5] we aimed to detect potential genomic signatures of domestication in two

independent pairs of wild/domesticated Atlantic salmon populations of Canadian and Scottish origins, respectively. Our results suggest that genetic drift may have override the effect of artificial selection and/or point toward a different genetic basis underlying the expression of similar traits in different domesticated strains. Finally, it is likely that domestication may predominantly target polygenic traits (e.g., growth) such that its genomic impact might be more difficult to detect with methods assuming selective sweeps.

A second application from the period concerns the development of an SNP-chip for genetic breeding of Tilapia. In [6] we describe all the SNP-chip development and application. Nile tilapia (*Oreochromis niloticus*) is one of the most cultivated and economically important species in world aquaculture. Males reach market size before females and reproduction during growth is a major problem that generates heterogeneous sizes of fish at harvest. For these reasons, identifying genomic regions associated with sex determination is a research topic of great interest. The objective of this study was to identify genomic variants associated with sex determination in three commercial populations of Nile tilapia. We detected a genome-wide significant signal comprising 36 SNPs, located on chromosome 23 spanning a genomic region of 536 kb. Ten significant SNPs intercept the anti-Müllerian hormone gene. Other significant SNPs were located in the neighboring Amh gene region. This gene has been strongly associated with sex determination in several vertebrate species, playing an essential role in the differentiation of male and female reproductive tissue in early stages of development. This finding provides useful information to better understand the mechanisms underlying sex determination in Nile tilapia.

References (5.b.-5.d.)

1. M. Aite, M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, N. Loira, G. Carrier, O. Dameron, N. Guillaudeux, M. Latorre, S. Mendoza, G. V. Markov, A. Maass, A. Siegel, Traceability, reproducibility and wiki-exploration for "`a-la-carte" reconstructions of genome-scale metabolic models. Plos Computational Biology 14 (5) (2018) e1006146.
2. Dinka Mandakovic, Ángela Cintolesi, Jonathan Maldonado, María Paz Cortés, Sebastián Mendoza, Méziane Aïte, Alexis Gaete, Francisco Saitua, Verónica Cambiazo, Anne Siegel, Alejandro Maass, Mauricio González, Mauricio Latorre, Comparing two genome-scale metabolic models of microbacterium species isolated from the Atacama Desert. Submitted 2018.
3. P.A. Saa, M. P. Cortés, D. Bustos, J. López, A. Maass, E. Agosín, Expanding metabolic capabilities using novel pathway designs: computational tools and applications. Submitted 2018.
4. D. Travisany, E. Goles, M. Latorre, M.-P. Cortés, A. Maass, Generation and robust- ness of boolean networks to model Clostridium difficile infection. Accepted Journal of Natural Computing 2019.
5. M.E. López, L. Benestan, J.S. Moore, C. Perrier, J. Gilbey, A. Di Genova, A. Maass, R. Neira, J.P. Lhorente, K. Correa, D. Díaz, L. Bernatchez, J.M. Yáñez, Independent and parallel genomic signatures of selection underlying domestication in two Atlantic salmon (Salmo salar L.) populations. Evolutionary Applications (12) (2019) 137-156.
6. Giovanna Cáceres, María E. López, María I. Cadiz, Grazyella M. Yoshida, Ana Jedlicki, Ricardo Palma, Dante Travisany, Diego Díaz, Alejandro Maass, Jean P. Lhorente3, José M. Yáñez, Fine mapping using whole-genome sequencing confirms anti-Müllerian hormone as a major gene for sex determination in farmed Nile tilapia (Orechromis niloticus). Submitted 2018.

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.

Overall, the evaluation by the panel of peers has been very positive and we thank them for their candor and support. Both of them are aware of the limitaitons imposed by the funding the Center receives and are adamant about the program providing, at the very least, continuity to the project. Their criticism and observations are mostly of a strategic nature, or, how to make the most of what we have for the maximum impact. A few comments on the major points they raise:

> **Reviewer 1:** *My opinion is that, in the next two years, one can expect more publications in journals with an impact factor >7, and therefore garnering much international recognition in a second funding period.*

> **Reviewer 2:** *The center's impact is fine, although currently not growing. Growth is difficult, though, for the lack of increasing funding. In order to convince the international research community and thus indirectly the national funding sources and politics that permanent and possibly more funding is required and would be well spent, the center should strive for landmark publications above an impact factor of 10. This reviewer is very well aware of the difficulties connected to this, including the partially difficult and biased reviewing process of some of the high-impact journals. Sometimes, it is scientifically a waste of time and resources to aim at publications in such journals. However, it is politically important for the center to get such papers published since a single one will create more attention than a lot of solid but less hyped publications. Therefore, the effort should be made.*

The matter of high impact papers is one of concern to us and has been the subject of debate. We have not had success putting our "Chilean made" papers into journals of the caliber of *Nature* or *Science*. Our papers in these journals have been as collaborators on a large international effort, such as the salmon, grape or potato genomes. We have submitted work aiming at high impact journals; for instance, the Huilliche genomes went through long review cycles in two top journals only to be rejected in the end for unclear (possibly unfair) reasons. Despite the above, we think it is only a matter of time until we get to the stage where our work will be showcased in one of the major journals; as the reviewer says, this may be the only way to draw attantion to the science, good as it may be. Finally, we can say that the average impact of our work (5.0), a level that has been sustained throughout the project, while not outstanding is quite respectable and indicates that we can compete in the "big leagues".

> **Reviewer 2:** *The number of postdocs and PhD students shrunk further in 2017, although their number is fine nevertheless. Every effort should be made to get young scientists to the center. This should not be at a cost to the science done there, however.*

As we explained above, part of the reason for the decline in number of trainees can be attributed to the impending end of the funding period, which, especially in the case of postdocs, is a detriment to our capacity to hire. These diminished from 28 in

2017 to 23 in 2018. An additional factor may be that the Director was on sabbatical leave for one year (2016-2017) and a PI, Dr. Gutiérrez, was also on sabbatical for half of 2018. Nonetheless, the number of students sharply increased this past year, from 43 in 2017 to 61 in 2018. Importantly, several of our ex-postdocs have obtained permanent positions. To name a few alumni that now have academic positions, Elena Vidal and Leonardo Valdivia at Universidad Mayor, Mauricio Latorre at Universidad de O'Higgins and Rodrigo Assar and Nicolás Tobar at the Universidad de Chile.

Reviewer 2: *The center may consider teaming up with other, similar activities in South-America in order to form an international, South-American network.*

The current state of science in South America is not as auspicious as we would like. Most of our neighboring countries are facing economic hardships and reduction in their gevernments' support for science. We have many excellent colleagues in the region and we collaborate with many of them. However, we have not seen initiatives similar to the FONDAP program where we could establish more institutional relationships. In the case of the 1000 genomes project, we have been able to come into contact with groups in Argentina, Brazil and Colombia, countries where there is a strong scientific base and a longstanding tradition in research. We are currently discussing ways in which to integrate our work into a Latin American framework for the protection of natural diversity and the right to have a "first look" at the genomic heritage of our continent.

Reviewer 2: *For future evaluations, it would be good, if the publications could be ordered according to the subprojects (Aims) from which they originate.*

As stated above, our aims are highly synergistic and published articles will often belong to more than one aim. In any case, articles are already declared in the reported tables as belonging to a specific aim. We refer the reader to any of the sections in the table and to examine the column "Sub-Program(s) or Research Line(s)" where we indicate a number from 1 to 5 that corresponds to the specific aim of the project. This is valid not only for articles but for postdocs, students, congress presentations, etc.

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.

Outreach and Media appearances.

1000 Genomes Project.

A large fraction of our outreach efforts and, consequently, our media impact, was related in 2018 to our leading role in the Chilean 1000 Genomes Project (a separate report on all of these efforts is supplied in Annex 2). To launch the initiative, we decided to announce a nationwide contest for high school students that would involve a sequencing experiment in their own classroom. To our knowledge, this is absolutely unprecedented and the announcement was received with a high degree of interest by potential participants. We had collaboration from the outreach office at CONICYT, EXPLORA, which helped us disseminate the announcement in all regions of Chile. In parallel, we created a web site (www.1000genomas.cl) that was used to promote the competition. A total of 65 applications were received for 10 selected groups. Each team, consisting of a maximum of 10 students belonging to a high school, had to submit an essay expressing their opinion on the importance of genomics to Chile. We asked for permission from the School Director for the activity, required parental consent (in writing) and requested that the teacher in charge had experience with extracurricular scientific activities. A panel of five judges, one selected by each Center of Excellence, was tasked with evaluating the essays and also to ensure adequate representation of public schools, regional institutions and gender balance. The ten selected schools were notified in June and were advised the activity would take place in all ten cities simultaneously in late August. The schools were widely distributed (over 3,000Km) throughout the country (only 2 were from the Santiago Metropolitan Region) and had a majority of public nature (5 completely public, 3 subsidized and 2 private). We next decided on the type of experiment to be carried out in the schools. As we wanted to make the experiment relatable to the students and of biological and scientific relevance, we sought to work with an organism

The group of students from the school in Toconao (II Region) that participated in the sequencing experiment.

The call for applications

that would be ubiquitous, easy to find and where no genomic information was yet available. We settled on the common pill bug (most common species are of the *Armadillium* genus) even though there are several species in Chile. In this way, the students could have similar projects and would be able to share the results and compare them. Further, we stated that, if new data was generated, the students would all share authorship on any scientific communication generated from this data. Permission form the Chilean Agriculture authorities (*Servicio Agrícola y Ganadero*) to collect the specimens was obtained. To sequence in the schools, we took advantage of a generous gift of ten MinIon sequencers and reagents, from Oxford Nanopore Technologies (UK), equipment that can be easily deplyoed in any environment for *in situ* sequencing of DNA or RNA. To use these sequencers away from the lab, we purchased a set of 11 portable computers with the capacity to run the sequencing experiment and store the data. We also acquired 10 sets of molecular biology reagents and suppplies in order to have the appropriate material at each sequencing site. All of the students were given materials such as lab coat, etc. The school students were prepared in advance so that they understood the nature of the experiment and the theoretical background of what was to be done. Twenty five graduate students were trained as monitors and were assigned, in pairs, to visit each of the schools on the designated days of the experiment. The visits took place on the week of August 20, 2018 and the monitors went to their assigned schools to carry out the experiment; sequencing was carried out over two days, Aug 23-24. At the end of the experiment (Aug. 24), all of the participating schools connected via video conference to Santiago where we held a closing ceremony together with the official launch of the 1000 Chilean Genomes Project. Invited authorities were shown the result and were able to see and hear the reporting from each school, some in very remote locations of the country. The CGR Director led the ceremony and also gave a speech to communicate the main points and rationale for the project. The entire event was also streamed on the web (2000 viewers were connected at one point) and has been uploaded permanently to YouTube (https://www.youtube.com/watch?v=96XdV8C-FRM).

As a result of the sequencing by the children, we obtained 12.8Gb of sequence data which is currently being processed for assembly of a genome.

Authorities that participated in the 1000 Genomes project launching. From left to right: Dr. Ennio Vivaldi, Rector of the University of Chile; Dr. Carolina Torrealba, Director of the Millennium Scientific Initiative; Dr. Christian González, Director of GERO; Dr. Miguel Allende, Director of the CGR; Dr. Alejandro Maass, Director of the CMM; Dr. Sergio Lavandero, Director of the ACCDis; Dr. Luis Larrondo, Director of iBio; Dr. Mario Hamuy, President of CONICYT.

Presentation of the 1000 Genomes Project to the Chilean Senate

Both the school activity and the 1000 Genomes effort itself were widely announced in the press, radio, print, web and television. Other repercussions also involved interviews with government and political leaders. Three CGR Researchers were invited to present the project before the Chilean Senate Committee on Science and Technology.

Left: Drs. Maass and Montecino during the presentation to the Senate Committee. Right: Dr. Allende explains the project to Senators Girardi, Coloma and Chahuán.

The hearings can be viewed entirely on the Senate's web site (select Dec 17th, 2018 was the date): http://tv.senado.cl/tvsenado/comisiones/permanentes/desafios-del-futuro-ciencia-tecnologia-e-innovacion/comision-desafios-del-futuro-ciencia-tecnologia-e/2018-12-17/170831.html

Media and Press Appearances

Among the press appearances of members of the CGR during 2018, besides those related to the 1000 Genomes Project, most had to do with our role as a valid opinion source for topics related to genome editing, stem cells, human genetics and science policy. We list here some of these instances with links.

1. Tele13 (television channel) 5 Nov 2018. *U. de Chile participará en proyecto que busca secuenciar genéticamente toda la vida del planeta.* http://www.t13.cl/noticia/tendencias/u-chile-participara-proyecto-busca-secuenciar-geneticamente-toda-vida-del-planeta

2. T13 Radio (FM radio) 6 Nov 2018. *Es importante para la salud pública tener la visión de los genomas humanos*. https://www.tele13radio.cl/?id_podcast=120181113115935

3. Duna Radio (FM radio). 13 Sep 2018. Interview of the CGR Director on the program "Aire Fresco": *Proyecto para secuenciar el genoma de mil chilenos y crear un repositorio nacional: "Va a haber una gran demanda por estas tecnologías".* http://www.duna.cl/programa/aire-fresco/2018/09/13/proyecto-para-secuenciar-el-genoma-de-mil-chilenos-y-crear-un-repositorio-nacional-va-a-haber-una-gran-demanda-por-estas-tecnologias/

4. El Mercurio (newspaper/web). 3 Apr 2018. *Científicos del país crearán un catálogo genético de Chile*. http://www.economiaynegocios.cl/noticias/noticias.asp?id=456845

5. Canal 2 (TV station in San Antonio, Chile). 26 Apr 2018. Interview with the CGR Director. https://www.youtube.com/watch?v=0PcL88ePqMk

6. CNN Chile. (television channel; interview) 30 Nov 2018. *Científicos de todo el mundo cuestionan la metodología detrás de los primeros bebés humanos genéticamente modificables*. https://www.futuro360.com/videos/cientificos-de-todo-el-mundo-cuestionan-la-existencia-de-los-primeros-bebes-humanos-geneticamente-modificables_20181130/

7. Las Ultimas Noticias (newspaper/web). 18 Apr 2018. Escolares buscarán saber que hay en el ADN de este bicho. http://www.lun.com/Pages/NewsDetail.aspx?dt=2018-04-18&PaginaId=12&bodyid=0

8. Revista Capital (magazine/journal) 10 May 2018. ¿De que estamos hechos? https://www.capital.cl/de-que-estamos-hechos/

9. Que Pasa (magazine/journal) 22 Oct 2018. El Código Genético Chileno. https://www.latercera.com/que-pasa/noticia/codigo-genetico-chileno/304889/

10. El Mostrador (electronic newspaper). 4 Dec 2018. *Científicos chilenos se declaran en alerta ante caso de gemelas modificadas genéticamente en China*. https://www.elmostrador.cl/cultura/2018/12/04/cientificos-chilenos-se-declaran-en-alerta-ante-caso-de-gemelas-modificadas-geneticamente-en-china/

11. CNN Chile (television channel; interview) 16 Nov 2018. *Un "Arca de Noé" en el siglo XXI: El ambicioso plan para registrar el ADN de todas las especies del mundo*. https://www.futuro360.com/videos/un-arca-de-noe-en-el-siglo-xxi-el-ambicioso-plan-para-registrar-el-adn-de-todas-las-especies-del-mundo_20181116/

12. CNN Chile (television channel; interview) 29 Aug 2018. *El ambicioso proyecto que busca descifrar el genoma de los chilenos*. https://www.cnnchile.com/lodijeronencnn/el-ambicioso-proyecto-que-busca-descifrar-el-genoma-de-los-chilenos_20180829/

13. UChile Noticias (university web site). 27 Aug 2018. *Autoridades universitarias y de gobierno participaron en el lanzamiento del "Proyecto 1000 Genomas".* http://www.uchile.cl/noticias/146455/se-realiza-lanzamiento-del-proyecto-1000-genomas

14. ADN Radio 91.7 (FM radio). 27 Aug 2018. *Genetista y los chilenos: "Seríamos un 50% europeo, un 45% nativo americano, y el 3% africano"*. http://www.adnradio.cl/noticias/sociedad/genetista-y-los-chilenos-seriamos-un-50-europeo-un-45-nativo-americano-y-el-3-africano/20180827/nota/3791627.aspx?ssm=whatsapp

15. Diario El Sur de Concepción (regional newspaper). 9 Apr 2018. *El proyecto que acercará a Chile a la medicina personalizada*. http://www.elsur.cl/impresa/2018/04/09/full/cuerpo-principal/16/

16. Radio Universidad de Chile 102.5 (radio/web). 6 Nov 2018. *U. de Chile integra proyecto para secuenciar genéticamente toda la vida del planeta*. https://radio.uchile.cl/2018/11/06/u-de-chile-integra-proyecto-para-secuenciar-geneticamente-toda-la-vida-del-planeta/

17. El Dínamo (web news service). 6 Nov 2018. *U. de Chile será parte de proyecto que busca secuenciar genéticamente toda la vida del planeta*. https://www.eldinamo.cl/educacion/2018/11/06/u-de-chile-sera-parte-de-proyecto-que-busca-secuenciar-geneticamente-toda-la-vida-del-planeta/

18. AdPrensa (news service). 5 Nov 2018. *U. de Chile será parte de proyecto que busca secuenciar genéticamente toda la vida del planeta*. https://www.adprensa.cl/actividades-academicas/u-de-chile-sera-parte-de-proyecto-que-busca-secuenciar-geneticamente-toda-la-vida-del-planeta/

19. Las Ultimas Noticias (Newspaper/web). 7 Nov 2018. *Biólogo explica el fabuloso proyecto que secuenciará el ADN de 1.5 millones de especies*. http://www.lun.com/Pages/NewsDetail.aspx?dt=2018-11-07&NewsID=413408&BodyID=0&PaginaId=2

20. Hoy por Hoy (newspaper/web). 7 Nov 2018. *U de Chile participará en proyecto que secuencia ADN*. http://www.hoyxhoy.cl/2018/11/07/full/cuerpo-principal/6/

21. Cooperativa 93.3 (FM Radio/web). 16 Aug 2018. *¿En que consiste la edición genómica?* https://www.cooperativa.cl/noticias/sociedad/salud/en-que-consiste-la-edicion-genomica/2017-08-16/112058.html

22. Radio Pauta (FM radio/web). 12 Apr 2018. Interview with the Director of the CGR on the program "La Invención del Futuro". http://www.pauta.cl/programas/la-invencion-del-futuro/la-invencion-del-futuro-12-de-abril

23. Sociedad de Biología de Chile (web/newsletter) 26 Apr 2018. *Dr. Miguel Allende: 1000 Genomas, el desafío genómico de Chile*. https://www.biologiachile.cl/2018/04/26/dr-miguel-allende-1000-genomas-el-desafio-genomico-de-chile/

24. T13 Radio (radio/web). 4 Sep 2018. *"1.000 Genomas Chile": Científicos buscan descifrar el código genético chileno*. http://www.t13.cl/noticia/nacional/1.000-genomas-chile-cientificos-buscan-descifrar-codigo-genetico-chileno.

25. Radio Universo 93.7 (radio/web). 5 Sep 2018. Interview with the CGR Director. https://www.universo.cl/miguel-allende-coordinador-del-proyecto-1000-genomas-se-viene-la-medicina-personalizada-donde-cada-persona-tendra-su-genoma-en-su-ficha-clinica-para-que-se-disene-una-respuesta-medica-especifica-para-ella/radio/2018-09-05/190303.html

26. Radio Cooperativa 93.3 (FM radio/web). 23 Apr 2018. *Interview of the CGR Director on the program "La Historia es Nuestra"; Cómo somos (genéticamente) en Chile*. http://www.cooperativa.cl/noticias/sociedad/ciencia/la-historia-es-nuestra-como-somos-geneticamente-en-chile/2018-04-23/164602.html

27.La Hora (newspaper/web). 30 Jul 2018. *Buscan apoyo para revelar la identidad genética de Chile*. http://www.lahora.cl/2017/07/buscan-apoyo-revelar-la-identidad-genetica-chile/

28. Xinhua News Agency (news service/web). 15 Apr 2018. *ESPECIAL: Chile conocerá por primera vez la genética de su población*. http://spanish.xinhuanet.com/2018-04/17/c_137115816.htm

29. La Tercera (newspaper/web). 18 Apr 2018. *Estudiantes chilenos podrán secuenciar genéticamente por primera vez al chanchito de tierra*. https://www.latercera.com/tendencias/noticia/estudiantes-chilenos-podran-secuenciar-geneticamente-primera-vez-al-chanchito-tierra/137209/

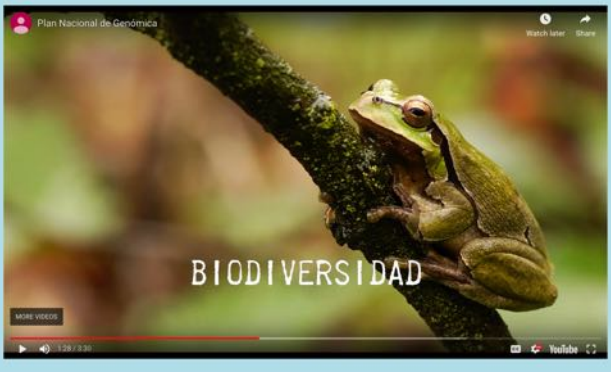
Special TV program on the CGR's Research

We hired a specialized communications company to produce and air a half hour program that became part of the series "*Exploradores: del átomo al cosmos*". For this episode, CGR Principal Investigators were interviewd and the main areas of research were highlighted. It is noteworthy to point out that the program showed footage taken by the CGR team during their field work in the Atacama Desert, as well as images from the work carried out in CGR labs. This program is shown nationwide through the channel *24 Horas* and the three recorded chapters shown on the program are placed on YouTube permanently. They can be seen here: https://youtu.be/cCYIk2mAQjo; https://youtu.be/a2RY1RBt7KU; https://youtu.be/Qu-IMFxydLA

A scene captured from the TV program dedicated to the CGR's research.

Ideas Conciencia

Towards the end of 2018, the Chilean Congress launched a competition for ideas to stimulate science and scientific literacy among the population; it was named *Ideas Conciencia*. To participate, interested organizations of any kind had to submit a written summary and a short (<3min) video of the topic they wished to promote. The CGR and the 1000 Genomes initiative decided to participate and prepared a short video combining animation and stock images aimed at highlighting the importance of genome sciences for our country. Among the hundreds of entries, our video was ranked among the top 20 and very positive comments were received. The video and proposal are displayed permanently on the Ideas Conciencia web site (https://www.ideasconciencia.cl/proyectos/5b9d9b65a7bed673de45481d) and also on YouTube: https://youtu.be/LyAk6Ri-uRE



Plan Nacional de Genómica

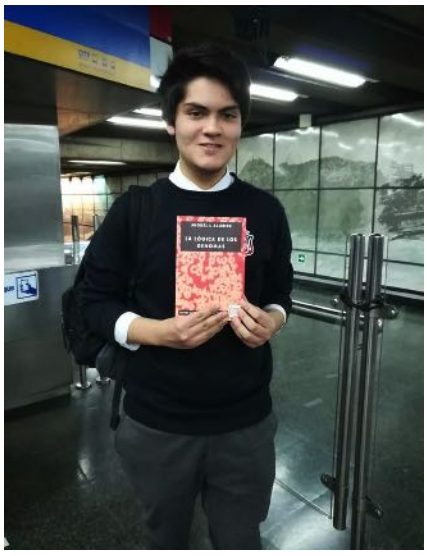
Futuro Ahora

Dr. Allende and Dr. Florencia Tevy were interviewed for the Futuro Ahora series sponsored by the Congreso del Futuro, Channel T13 and El Mostrador. The program can be seen at the following web site: https://www.elmostrador.cl/noticias/pais/2018/11/17/cientificos-llaman-a-regular-la-informacion-genetica-en-chile-para-evitar-el-uso-perverso-de-nuevos-tipos-de-discriminacion/

Book publication

Dr. Allende published a book for the general public entitled "*La Lógica de los Genomas*". The book is a simplified description of the field of genomics and its econonomical, social and ethical implications. Thus far, 600 copies have been printed and distributed for free among school students, school teachers and other citizens that have expressed interest, all over the country. It was also sent to opinion leaders, members of academia and the government, including the President of Chile, whom expressed his thanks for the gift. The book is available electronically in PDF, Kindle and E-Reader formats. It can be freely downloaded here: http://www.genomacrg.cl/la-logica-de-los-genomas.html

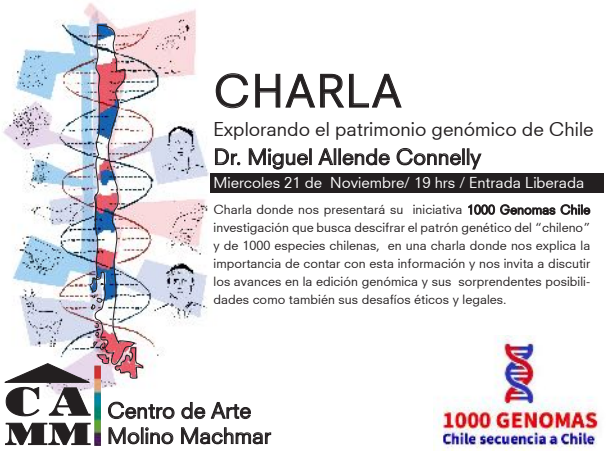
Copies of the book were given away at a science fair in Antofagasta.



Three copies were awarded to participants of an online contest.

Outreach activities and interaction with society

1. Dr. Allende gave a lecture for the citizens of Puerto Varas entitled "Explorando el patrimonio genético de Chile" on Nov. 21, 2018. About 60 people attended.



2. Science Fair "Ciencia en el parque". The CGR and the 1000 Genomes project had a stand at the open science fair organized by EXPLORA-CONICYT as part of the XXIV National Science and Technology week. Thousands of people attended the event which was held in the O'Higgins Park on October 6th, 2018.

Students at the Science Fair present their projects.

3. For the fifth consecutive year, we were in charge of one of the modules of the Portable Labs Workshops that aims to train secondary school teachers for experiments in the classroom in different areas of biology. We held the workshop from January 8 to 12 and 18 teachers attended. Later in the year, we visited each of those schools with a team of graduate students and did the experiments with the students. Every group then carried out projects and presented them at a Science Fair held at the University of Chile on November 7, 2018.

4. Among several presentations at schools in Santiago, we highlight the one called "*Charla sobre el ADN y el genoma*" given by Dr. Ariel Orellana. It was held at the Liceo polivalente 71 Guillermo Feliú Cruz, Comuna de Estación Central, on 24 April, 2018. The presentation can be viewed at the following web site: https://www.facebook.com/tvescolar/videos/1167162246753886/

Awards and Prizes.

- Dr. Alejandro Maass was incorporated as a Corresponding Member of the Chilean Academy of Sciences on August 22, 2018.

- Dr. Gutiérrez received the prestigious Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation. It is the first time a Chilean receives the award and the 45,000€ were awarded for cooperation with the Max Planck Institute in Plant Biology, Munich.

- Dr. Allende received a funding award from EMBO for the workshop "Exploring Genomic Landscapes", an event that was held in San Pedro de Atacama, Chile, January 10-12, 2019.

- Dr. Gutiérrez received a funding award from EMBO for the workshop "Systems and Synthetic Biology" to be held in April 2019.

- CGR Principal Investigator Rodrigo Gutierrez published an article in Plant Physiology in 2017 which had the 2nd highest Altmetric score of all articles downloaded in 2018. Altmetrics is a measure of how much attention journal articles are receiving based on discussions on the web, research blogs, social media, Wikipedia, and in public policy documents and the like. The article is titled: "The Next Generation of Training for Arabidopsis Researchers: Bioinformatics and Quantitative Biology". http://www.plantphysiol.org/content/175/4/1499

- Dr. Allende gave a guest lecture at the Universidad de Antofagasta on the 1000 Genomes Project. November 9, 2018.

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.

The Program Office works very efficiently and is always available for providing assistance to the proper functioning of the Center. We have no recommendations to offer and we extend our appreciation to the staff for their excellent work. The review process has also been fair and the panel of experts that has carried out the evaluations is very helpful.

We do, however, have a comment with regards to the decision making process at the highest levels of CONICYT. It is unsettling, to say the least, that with less than two years until the end of our funding period, we still do not know whether there will be any possibility for continuity of the CGR. This uncertainty is highly detrimental to our work and planning. For instance, we cannot higher new postdocs or students to work on the CGR's projects at this stage as we do not know if we will have funding for them (or the possibility of such funding). Also, to set up collaborations or new avenues of research, it is critical to be able to project these plans into the future with a degree of confidence, which we don't have at this time. It is our hope that the new organization of Chilean science with a Ministry at the helm can solve this perennial problem of absence of foresight and leadership.

X. ANNEXES

1. **Advisory Board´s Report**
2. **Communications and outreach materials**
3. **Manuscript by Di Genova et al., submitted to GigaScience.**