

CURSO DE POSGRADO

ÓMICAS Y ANÁLISIS DE GRANDES BASES DE DATOS EN VEGETALES APLICADOS A LA GENÉTICA Y AL MEJORAMIENTO

Encuentros sincrónicos: Miércoles 11, 18, 25 de noviembre y 2 y 9 de diciembre

Horario: 16 a 18h

Cupo: 35 alumnos.

Carga horaria: 30 h

Gratuito para socios SAG.

Para no socios se cobra una matrícula de \$120.000 (ciento veinte mil pesos).

Los miembros RIABIA-pro tendrán un descuento de 25% (\$90.000).

CUERPO DOCENTE

Dr. Guillermo Pratta — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR), Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario.

Dr. Javier Pereira da Costa — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR), Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario.

Dr. Flavio Spetale — Centro Internacional Franco Argentino de Ciencias de la Información y de Sistemas (CIFASIS, CONICET–UNR), Universidad Nacional de Rosario.

Dra. Tatiana Vega — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario.

Dr. Paolo Cacchiarelli — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR), Cátedra de Fitopatología, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario.

Dra. Dana Vázquez — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR), Cátedra de Genética, Facultad de Ciencias Agrarias, Universidad Nacional de Rosario.

Dr. Nicolas Defarge — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR)

Dra. Daniela Liebsch — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR).

Dr. Gustavo Rodríguez — Instituto de Investigaciones en Ciencias Agrarias de Rosario (IICAR, CONICET–UNR), Facultad de Ciencias Agrarias, Universidad Nacional de Rosario.

Modalidad

El curso se desarrollará en modalidad virtual y comprenderá 5 encuentros sincrónicos de hasta 2 horas cada uno, complementados con actividades prácticas asincrónicas. Las clases abordarán los fundamentos de las principales tecnologías ómicas, su aplicación a la genética y el mejoramiento vegetal, y las estrategias de análisis de grandes conjuntos de datos. Se incluirán herramientas de análisis multivariado, como ACP y AFM, así como aproximaciones de aprendizaje automático e inteligencia artificial. Las actividades prácticas estarán orientadas al uso de bases de datos y recursos bioinformáticos, al análisis e interpretación de resultados y a la resolución de casos aplicados.

Encuentro 1. Genómica vegetal y variación genética

Docentes: Dres. Daniela Liebsch y Gustavo Rodríguez

Contenidos teóricos: Organización y características de los genomas vegetales. Tecnologías de secuenciación. Ensamblado y anotación de genomas. Identificación de SNP, InDel y variantes estructurales. Genotipado de alta densidad. Conceptos básicos de diversidad genética, QTL y GWAS. Principales bases de datos genómicas vegetales.

Actividad práctica asincrónica (carga horaria: 2.5 h)

Exploración de una región genómica en bases de datos públicas. Identificación de genes, variantes y anotaciones funcionales asociadas a un carácter de interés.

Encuentro 2. Transcriptómica y expresión génica

Docentes: Dres. Paolo Cacchiarelli y Nicolas Defarge

Contenidos teóricos: Fundamentos y diseño experimental de RNA-seq. Control de calidad y procesamiento de datos. Cuantificación y normalización de expresión génica. Expresión diferencial. PCA, clustering y visualización. Análisis funcional y enriquecimiento. Identificación de genes candidatos.

Actividad práctica asincrónica (carga horaria: 2.5 h)

Interpretación de un conjunto de resultados de RNA-seq: PCA, heatmap, volcano plot y análisis funcional de genes diferencialmente expresados.

Encuentro 3. Proteómica

Docentes: Dres. Tatiana Vega y Javier Pereira da Costa

Contenidos teóricos: Principios generales de proteómica. Espectrometría de masas. Proteómica dirigida y no dirigida. Identificación y cuantificación de proteínas. Procesamiento y normalización de datos. Análisis multivariado. Aplicaciones.

Actividad práctica asincrónica (carga horaria: 2.5 h)

Análisis e interpretación de un conjunto simplificado de datos proteómicos o metabolómicos mediante PCA, agrupamiento e identificación de variables discriminantes.

Encuentro 4. Análisis de grandes bases de datos biológicos

Docente: Dr. Guillermo R. Pratta

Contenidos teóricos: Características de los datos ómicos de alta dimensionalidad. Preparación, depuración, transformación, estandarización y normalización de datos. Análisis exploratorio y relaciones entre variables. Análisis de Componentes Principales (ACP/PCA). Análisis de conglomerados. Análisis Factorial Múltiple (AFM/MFA). Integración de grupos de variables. Visualización e interpretación biológica de resultados multivariados.

Actividad práctica asincrónica (carga horaria: 2.5 h)

Análisis exploratorio de una base de datos vegetal mediante ACP, clustering y AFM. Interpretación de individuos, variables, contribuciones y relaciones entre distintos bloques de información.

Encuentro 5. Machine learning aplicado a genética y mejoramiento vegetal

Docentes: Dres. Dana Vazquez y Flavio Spetale

Contenidos teóricos: Conceptos básicos de aprendizaje automático. Aprendizaje supervisado en Clasificación y Regresión. Árboles de decisión, Máquinas de Vectores Soporte, Random Forest y XGBoost. Entrenamiento, validación (ajuste de hiperparámetros) y evaluación de modelos (rendimiento). Métricas de evaluación. Sobreajuste, generalización e interpretación de la importancia de variables. Aplicaciones a la predicción de fenotipos y clasificación de genotipos.

Actividad práctica asincrónica (carga horaria: 2.5 h)

Aplicación e interpretación de un modelo sencillo de clasificación o regresión para predecir un carácter vegetal a partir de variables genómicas, ómicas o fenotípicas.

Evaluación final integradora (carga horaria: 7.5 h)

Cada estudiante deberá elaborar una propuesta breve de aplicación de los contenidos del curso a su objeto de estudio actual, ya sea una tesis de grado o posgrado, una ayudantía de investigación o una línea de trabajo en la que participe. La propuesta deberá plantear una pregunta o problema biológico concreto e integrar al menos dos tipos de datos o enfoques abordados durante el curso (genómicos, transcriptómicos, proteómicos, fenotípicos, análisis multivariado o machine learning), justificando su elección. Se deberá describir brevemente el material de estudio, los datos a utilizar o generar, la estrategia de análisis propuesta y los resultados esperados, destacando de

qué manera esta aproximación podría aportar a la comprensión del problema y/o al mejoramiento vegetal.

BIBLIOGRAFÍA:

Belich, Y. E., & Pratta, G. R. (2023). Estimation of genetic variance components and assessment of narrow and broad sense heritabilities for maize ear rot in RIL populations derived from three biparental crosses. *Brazilian Archives of Biology and Technology*, 66. <https://doi.org/10.1590/1678-4324-2023220838>

Belser, C., Istace, B., Denis, E., Dubarry, M., Baurens, F. C., Falentin, C., Genete, M., Berrabah, W., Chèvre, A. M., Delourme, R., Deniot, G., Denoeud, F., Duffé, P., Engelen, S., Lemainque, A., Manzanares-Dauleux, M., Martin, G., Morice, J., Noel, B., Vekemans, X., D'Hont, A., Rousseau-Gueutin, M., Barbe, V., & Aury, J. M. (2018). Chromosome-scale assemblies of plant genomes using nanopore long reads and optical maps. *Nature Plants*, 4, 879–887. <https://doi.org/10.1038/s41477-018-0289-4>

Bianchi, M. B., Gibbs, P., & Pratta, G. R. (2019). Biplot approach with self-incompatibility analyses of sib diallel cross data for *Lysimachia monelli* (Primulaceae). *Plant Biosystems*. <https://doi.org/10.1080/11263504.2019.1587540>

Cabodevila, V. G., Cambiaso, V., Rodríguez, G. R., Picardi, L. A., Pratta, G. R., Capel, C., Lozano, R., & Capel, J. (2021). A segregating F2 population from a tomato second cycle hybrid allows the identification of novel QTL for fruit quality traits. *Euphytica*, 217(6). <https://doi.org/10.1007/s10681-020-02731-6>

Cacchiarelli, P., Goytia Bertero, V., Garcia Labari, I., Arce, D. P., Tapia, E., & Pratta, G. R. (2026). Association between molecular and quantitative variability in a tomato broad cross identified by sequence-specific markers developed from transcriptomic polymorphism for the small Heat Shock Protein (sHSP) cluster in chromosome six. *Genetic Resources and Crop Evolution*, 73, 85. <https://doi.org/10.1007/s10722-025-02672-w>

Cambiaso, V., Pratta, G. R., Pereira da Costa, J. H., Zorzoli, R., Francis, D. M., & Rodríguez, G. R. (2019). Whole genome re-sequencing analysis of two tomato genotypes for polymorphism insight in cloned genes and a genetic map construction. *Scientia Horticulturae*, 247, 58–66.

Cuba Amarilla, M. M., Peruzzo, A. M., Pioli, R. N., Hernández, F. E., Balaban, D. M., & Pratta, G. R. (2026). Resistance genes to *Diaporthe phaseolorum* var. *caulivora* in *Glycine max*: Qualitative and quantitative approaches. *Agronomía Mesoamericana*, 37, jdr4y941. <https://doi.org/10.15517/jdr4y941>

Del Medico, A. P., Cabodevila, V. G., Vitelleschi, M. S., & Pratta, G. R. (2019). Multivariate estimate of heritability for quality traits in tomato by the multiple factor analysis. *Pesquisa Agropecuária Brasileira*, 54, e00064. <https://doi.org/10.1590/S1678-3921.pab2019.v54.00064>

Del Medico, A. P., Cabodevila, V. G., Vitelleschi, M. S., & Pratta, G. R. (2020). Characterization of tomato (*Solanum lycopersicum* L.) generations according to three-way data analysis. *Bragantia*. <https://doi.org/10.1590/1678-4499.20190047>

Del Medico, A. P., Tenaglia, G. C., Lavalle, A. L., Vitelleschi, M. S., & Pratta, G. R. (2021). Analysis of mixed data to select banana clones (*Musa* spp.) to be included in a germplasm bank. *Journal of Basic and Applied Genetics*, 32(1), 25–42.

Del Medico, A. P., Vitelleschi, M. S., Lavalle, A. L., Tenaglia, G. C., & Pratta, G. R. (2024). Agroecological aptitude of the Northeast of Formosa (Argentinean subtropical region) for banana production assessed by multiple factor analysis. *Agriculture*, 14(11), 1904.

Di Giacomo, M., Vega, T. A., Cambiaso, V., Picardi, L. A., Rodríguez, G. R., & Pereira da Costa, J. H. (2023). An integrative transcriptomics and proteomics approach to identify putative genes underlying fruit ripening in tomato near-isogenic lines with long shelf life. *Plants*, 12, 2812.

Edwards, D., & Batley, J. (2010). Plant genome sequencing: Applications for crop improvement. *Plant Biotechnology Journal*, 8, 2–9. <https://doi.org/10.1111/j.1467-7652.2009.00459.x>

Ermini Starna, J. L., Tenaglia, G. C., & Pratta, G. R. (2024). Micropropagation of banana genetic resources adapted to the Argentine environment and generation of new variability for agronomic traits through somaclonal variation. *Genetic Resources and Crop Evolution*. <https://doi.org/10.1007/s10722-023-01840-0>

Faviere, G., Vitelleschi, M. S., & Pratta, G. R. (2024). Random Forest in plant genetics and breeding: Tomato as a model crop. *Journal of Basic and Applied Genetics*, 35(1), ART3.

Hernández, F. E., Peruzzo, A., Pratta, G. R., & Pioli, R. N. (2020). Polimorfismo molecular asociado a diversidad patogénica en interacciones de *Phomopsis longicolla* (teleomorfo desconocido) y *Glycine max* en Argentina. *Agrociencia*, 54(3), 313–326.

Lamas, C. Y., Urtasun, M. M., Giamminola, E. M., Pratta, G. R., Caruso, G. B., Morandini, M. N., & de Viana, M. L. (2022). Fruit and seed characterization of wild populations of an ancient Andean crop: *Solanum betaceum* (Cav.) (Solanaceae) in the Argentinian Yungas. *Genetic Resources and Crop Evolution*. <https://doi.org/10.1007/s10722-021-01223-3>

Lavilla, M., Fernández, A. E., & Pratta, G. R. (2026). Análisis morfogenético de aislamientos de *Cercospora kikuchii* provenientes de distintas regiones de la Argentina. *Agrosilvicultura y Medioambiente*, 4(1), 4–20. <https://doi.org/10.47230/agrosilvicultura.medioambiente.v4.n1.2026.4-20>

Li, Y., & Chen, L. (2014). Big biological data: Challenges and opportunities. *Genomics, Proteomics & Bioinformatics*, 12(5), 187–189. <https://doi.org/10.1016/j.gpb.2014.10.001>

Luna-Maldonado, A. I. (2026). Data-driven plant phenotyping: Integrating artificial intelligence and imaging technologies for precision crop improvement. In *Precision Agriculture – Data-Driven*

Technologies for Sustainable Food Production [Working Title]. IntechOpen.
<https://doi.org/10.5772/intechopen.1014679>

Pereira da Costa, J. H., Souza Canada, E. D., Ochogavía, A. C., Rodríguez, G. R., & Pratta, G. R. (2026). Genetic variability of gene expression in tomato fruits ripened on and off the vine: Cis-regulatory elements associated with differential transcription patterns in the most discrepant variety. *Plants*, 15(1), 53. <https://doi.org/10.3390/plants15010053>

Sun, Y., Shang, L., Zhu, Q.-H., Fan, L., & Guo, L. (2022). Twenty years of plant genome sequencing: Achievements and challenges. *Trends in Plant Science*, 27(4), 391–401. <https://doi.org/10.1016/j.tplants.2021.10.006>

Vazquez, D. V., Spetale, F. E., Nankar, A. N., Grozeva, S., & Rodríguez, G. R. (2024). Machine learning-based tomato fruit shape classification system. *Plants*, 13, 2357. <https://doi.org/10.3390/plants13172357>

Vitelleschi, M. S., & Pratta, G. R. (2024). Assessment of the genetic structure of two different tomato breeding populations by Principal Component Analysis. *FAVE – Sección Ciencias Agrarias*, 23, e0024.

Wang, W. P., Zhang, B., & Li, M. M. (2024). Big data analytics in biology: A systematic review of methods for large-scale data processing. *Computational Molecular Biology*, 14(3), 97–105. <https://doi.org/10.5376/cmb.2024.14.0012>