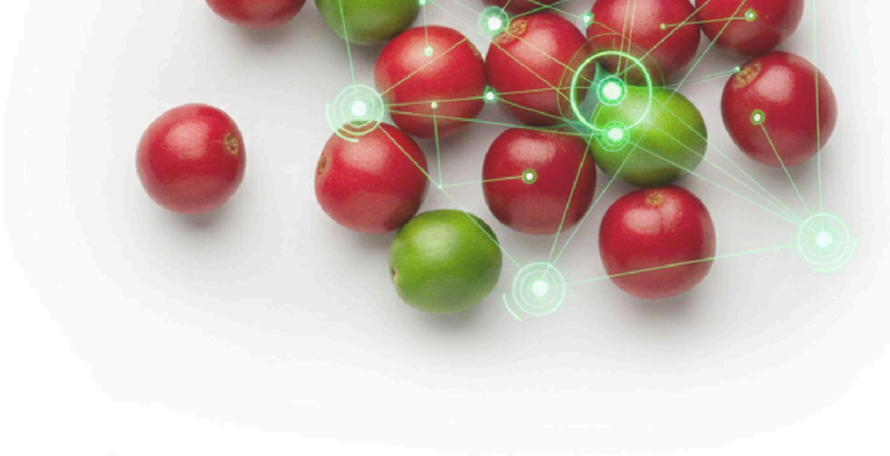




Bioinformática: Transformando Datos en Conocimiento Científico

Nicolás Ocampo Moreno
Pasante
Disciplina de Fitopatología

Cenicafé®
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia®

PARTE I



CUANDO LA BIOLOGÍA CAMBIÓ PARA SIEMPRE

La secuenciacion de ADN cambió las condiciones

El pasado



La biología antes de ADN: Observación física, descripciones manuales, datos limitados a lo que el ojo humano y el miscroscopio podían capturar.

El presente



El big data tecnológico: La Tecnología de Secuenciación de Nueva Generación (NGS) nos entregó millones de libros de información genética de la noche a la mañana.

El desafío de la biología moderna

Biología tradicional	Biología Moderna
Reto: Obtener suficientes datos en el campo o en el invernadero.	**Reto:** Darle sentido a terabytes de secuencias genéticas.
Herramientas: Pipetas, microscopio, bitacoras, etc.	**Herramientas:** Algoritmos, supercomputadoras, inteligencia artificial.
Cuello de botella: El tiempo de recolección en campo.	**Cuello de botella:** La capacidad de análisis computacional.

Biología + Computación + Matemáticas + Estadística =

Conocimiento científico

Entender el genoma del café es proteger el motor económico de cientos de miles de familias



Más que una bebida

El café es el sustento de la agricultura nacional y un patrimonio económico.

Defensa molecular

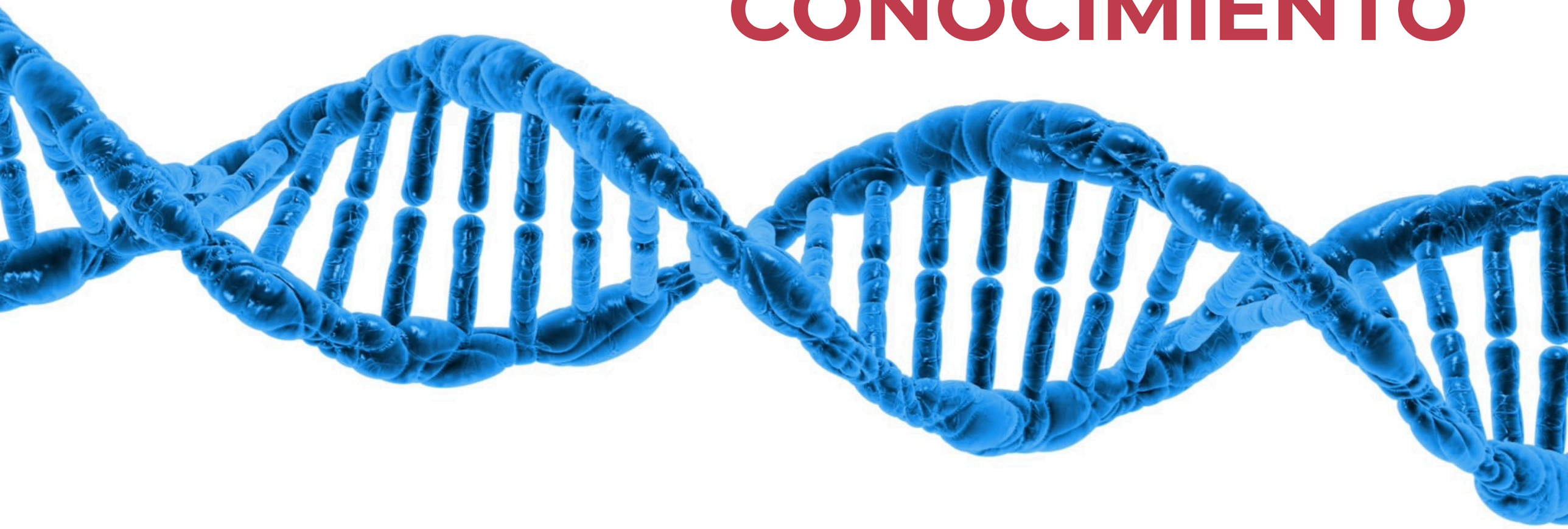
Para asegurar su calidad, resistencia a plagas y adaptación al clima, necesitamos entender su manual de instrucciones a nivel celular.

La visión de Cenicafé

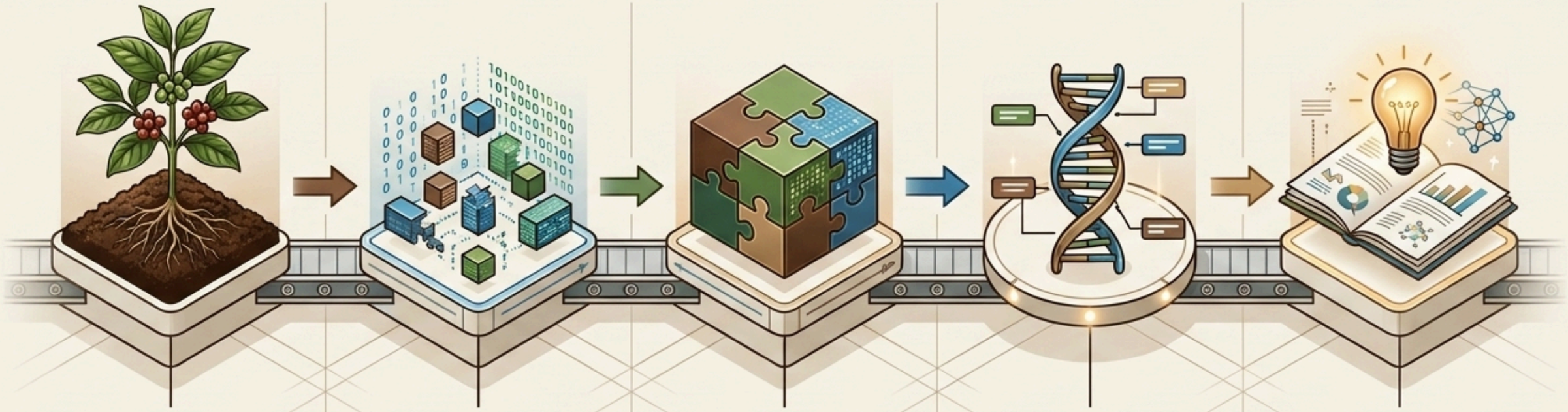
Liderar la comprensión genética de nuestros cultivos para garantizar el futuro de la caficultura colombiana.

PARTE II

DEL ADN AL CONOCIMIENTO



El flujo de trabajo bioinformático opera como una fábrica de ensamblaje de conocimiento



1.
Muestra Biológica:
La planta en el campo.

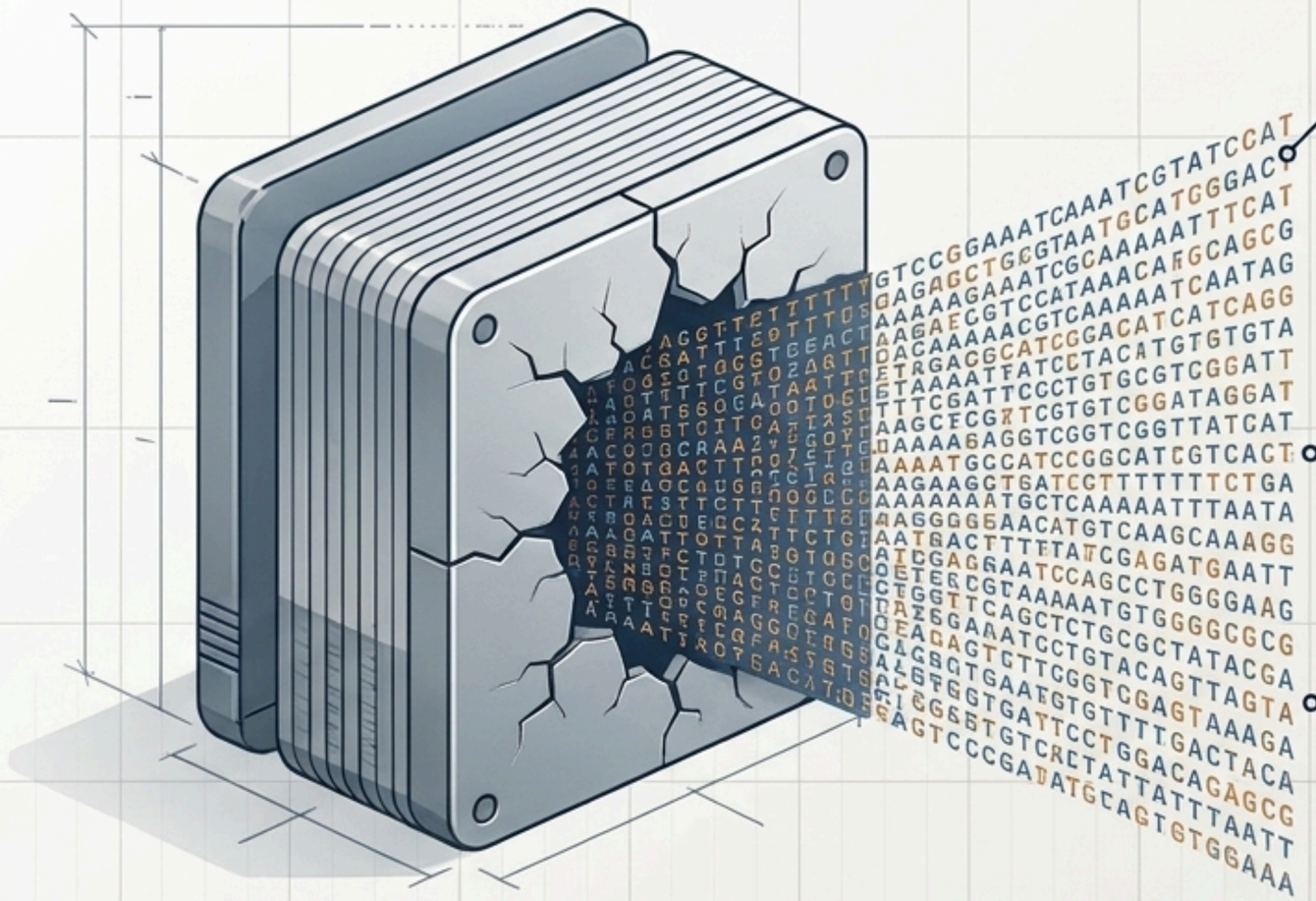
2.
Datos crudos (Materia prima):
Millones de piezas de información sueltas.

3.
Ensamblaje (Armando el rompecabezas):
Reconstrucción del genoma completo a partir de las piezas

4.
Anotación (Etiquetado):
Identificación de qué hace cada pieza

5.
Conocimiento Cientifico (Producto final):
Publicación, descubrimiento y aplicación.

El dato crudo: El registro fundamental



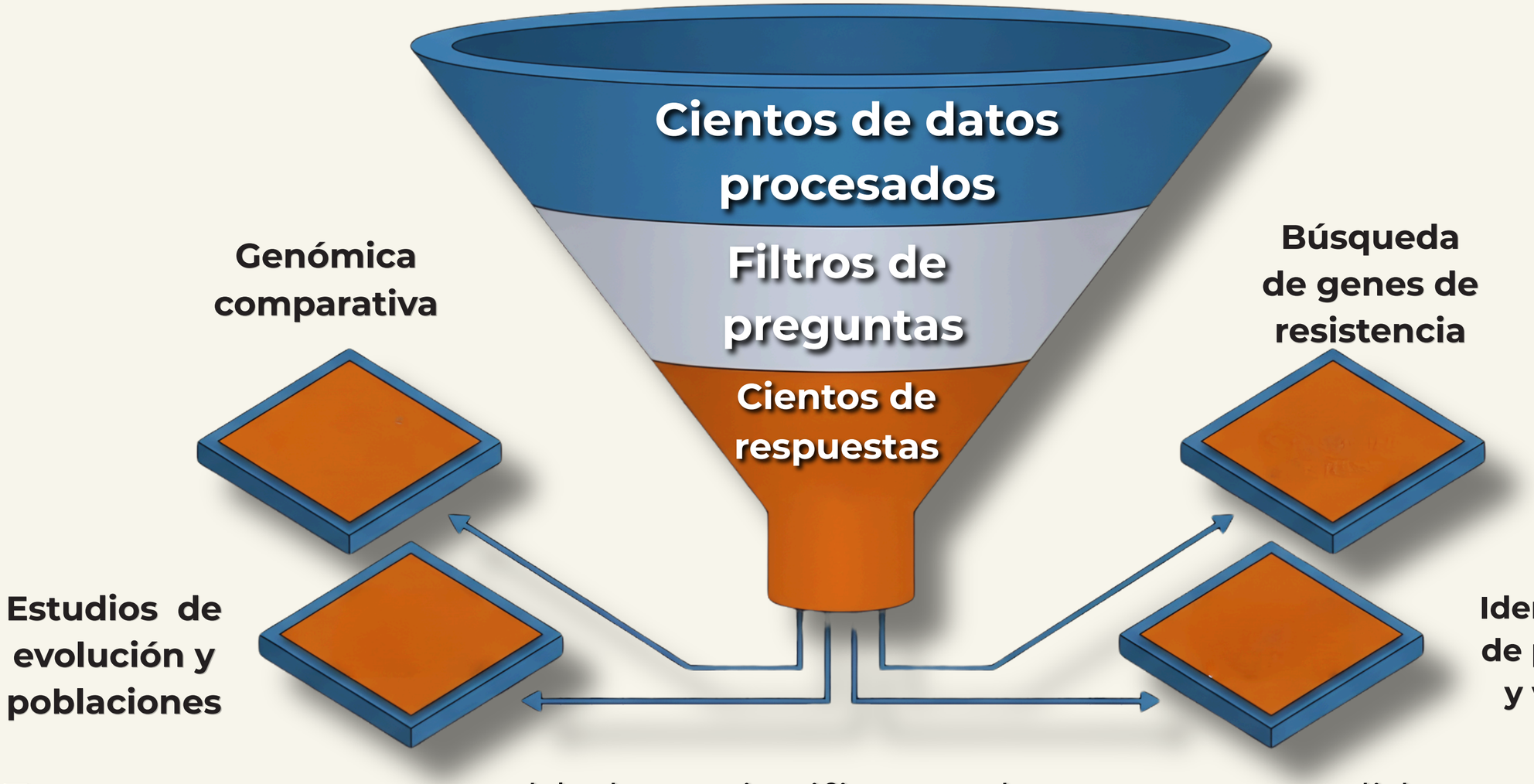
Millones de lecturas generadas por el instrumento

Aún no nos dicen qué gen está presente, ni qué función tiene

Archivos masivos en formatos especializados (FASTQ)

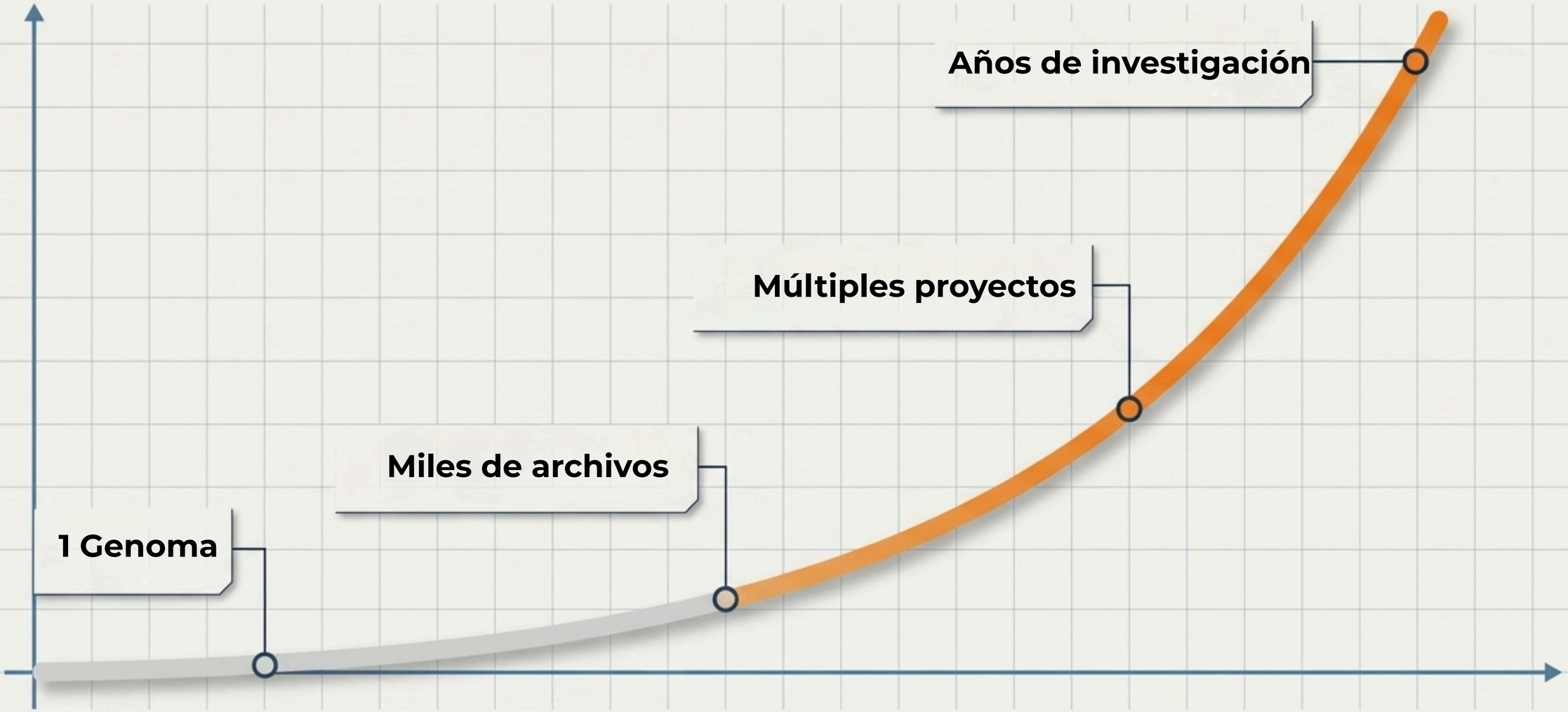
Si perdemos el dato crudo, perdemos parte de la capacidad de reconstruir el experimento.

La pregunta biológica



Tener un genoma ensamblado no significa que hayamos respondido una pregunta.

La escala cambió: De archivo a patrimonio



Lo que ayer era un archivo experimental en la computadora de un investigador, hoy se consolida como parte fundamental del patrimonio digital e intelectual de la institución.

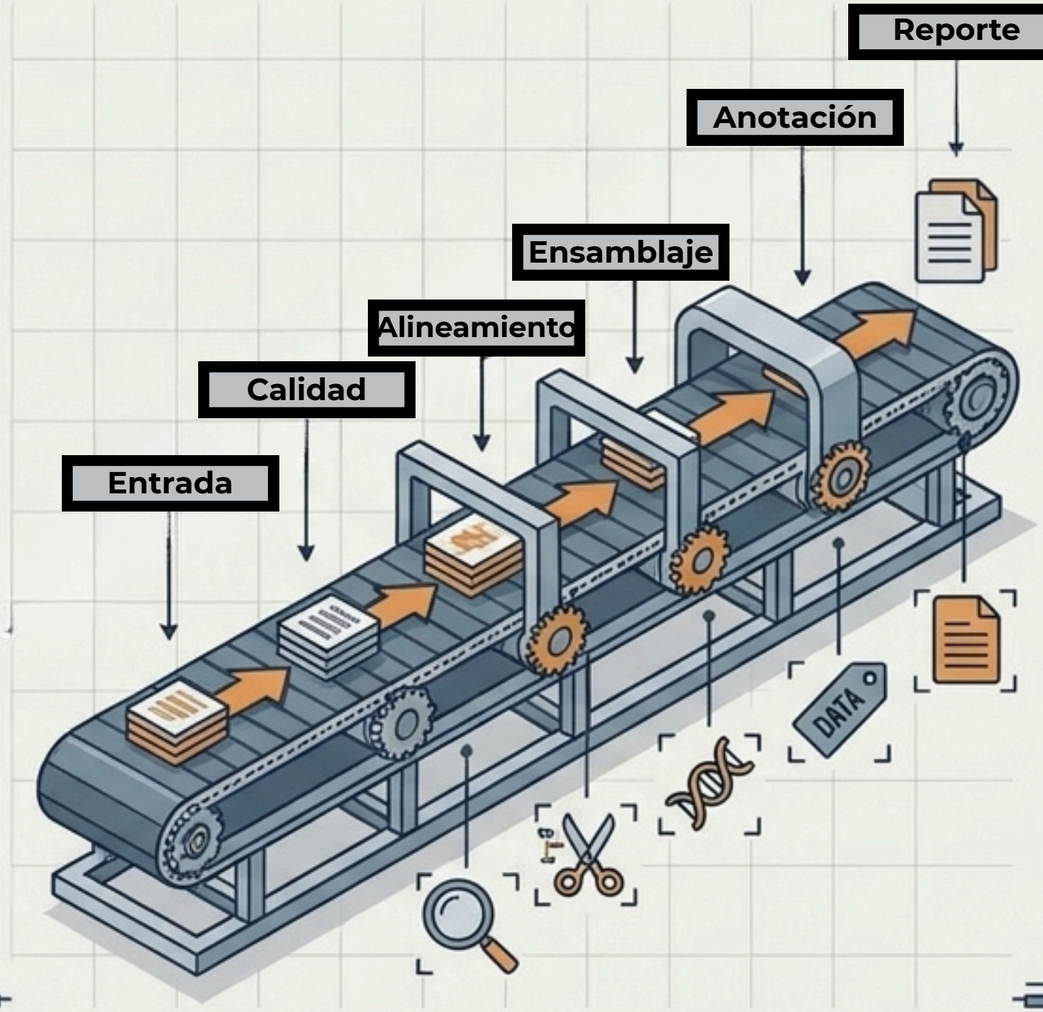
Cuando un análisis se convierte en un flujo de trabajo

Procesamiento manual



1 muestra = Viable.
10 muestras = Ineficiente.
Ciento de muestras = Imposible.

Flujo automatizado



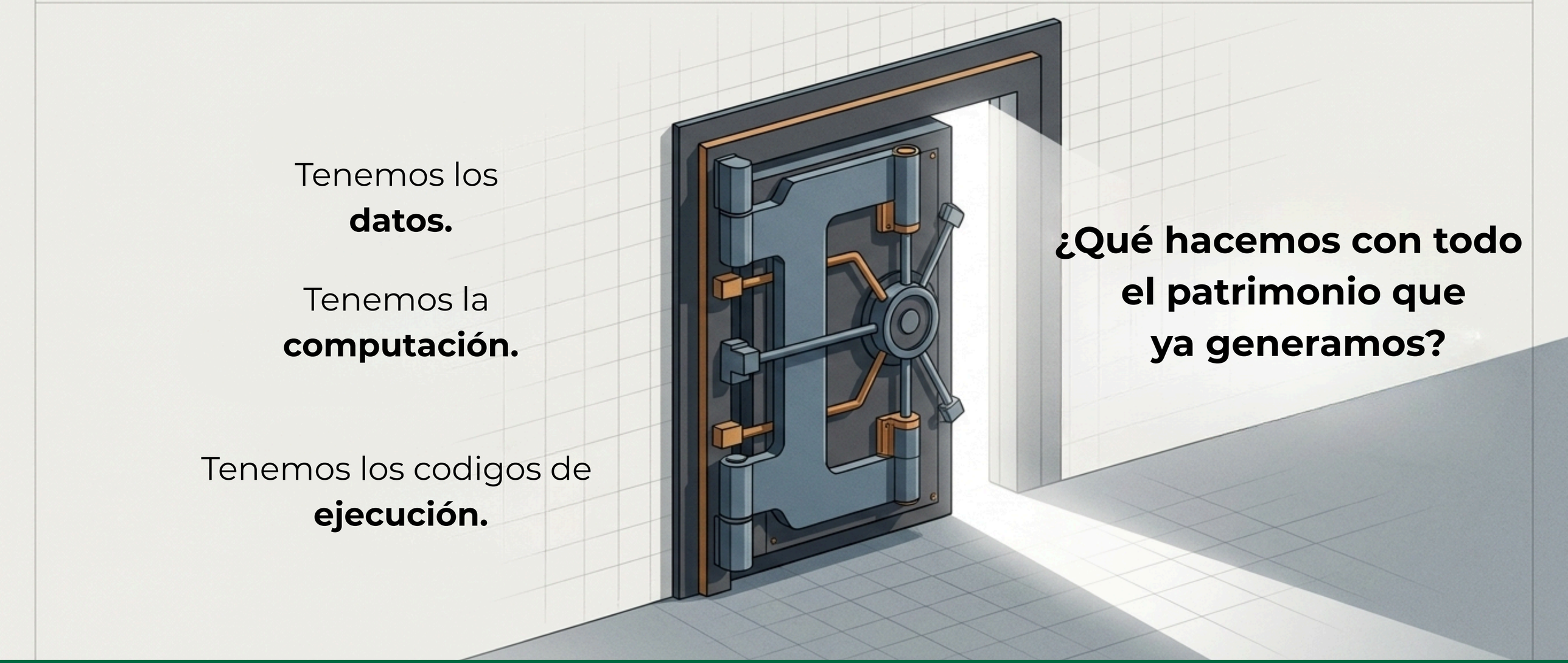
La automatización elimina el cuello de botella humano en el procesamiento de datos masivos.

La verdadera prueba

Tenemos los
datos.

Tenemos la
computación.

Tenemos los codigos de
ejecución.



¿Qué hacemos con todo el patrimonio que ya generamos?

PARTE III

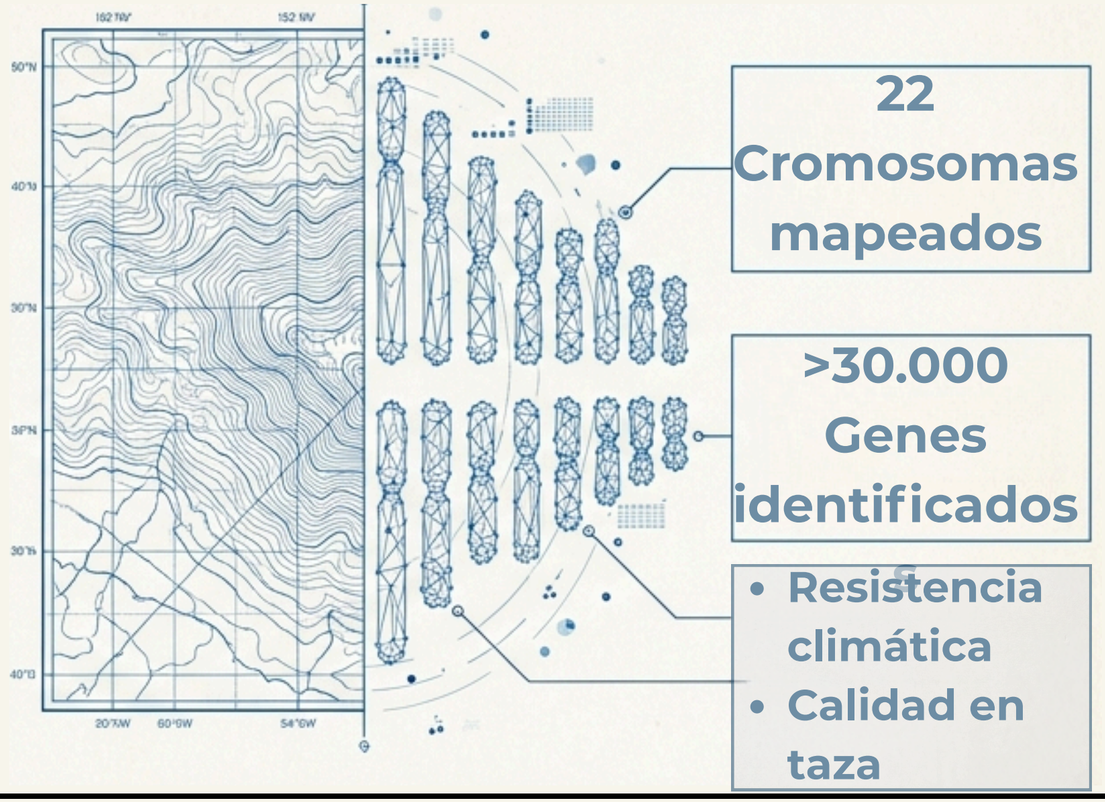
LA BIOINFORMÁTICA EN CENICAFÉ



SABEMOS
LO QUE HACEMOS

Cenicafé
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia

El pasado	El presente
	

Con este mapa genómico completo del café, podriamos identificar las coordenadas de la calidad en taza, el sabor y la resistencia climática.

El mapa genético del café

Coffea eugenioides

Aporta los planos base de calidad y perfil de taza.



Diploides



Coffea canephora

Aporta la robustez y resistencia al entorno.

Nació en Etiopía hace unos 500,000 años, creando una planta fértil de la cual desciende todo el café arábico actual.

Tetraploide



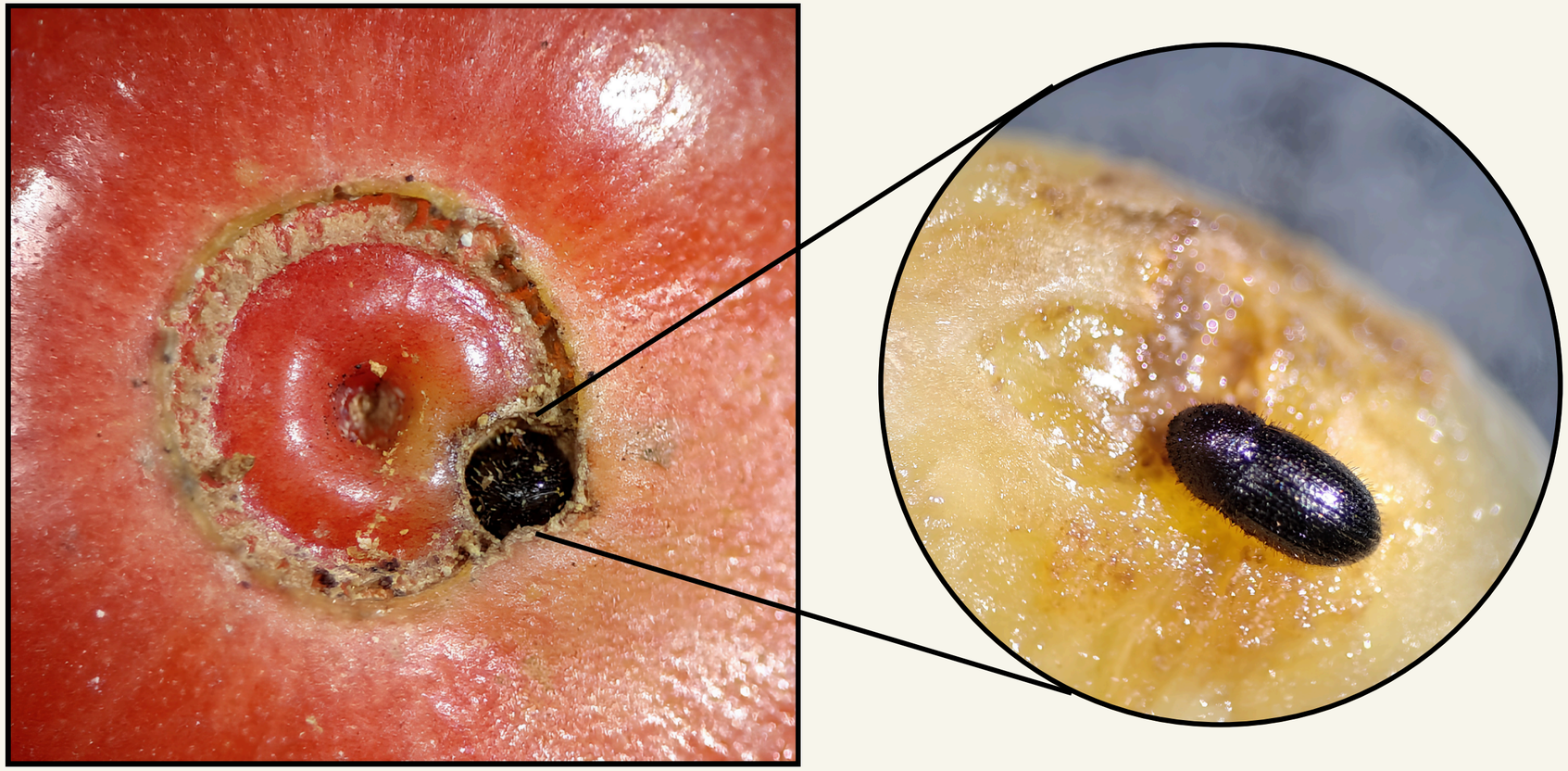
Coffea arabica

La síntesis entre calidad y resistencia en campo.

Para entender nuestra variedad, decodificamos a sus parentales evolutivos.

Este conocimiento nos permite identificar genes para enfrentar el cambio climático.

El enemigo invisible dentro del grano



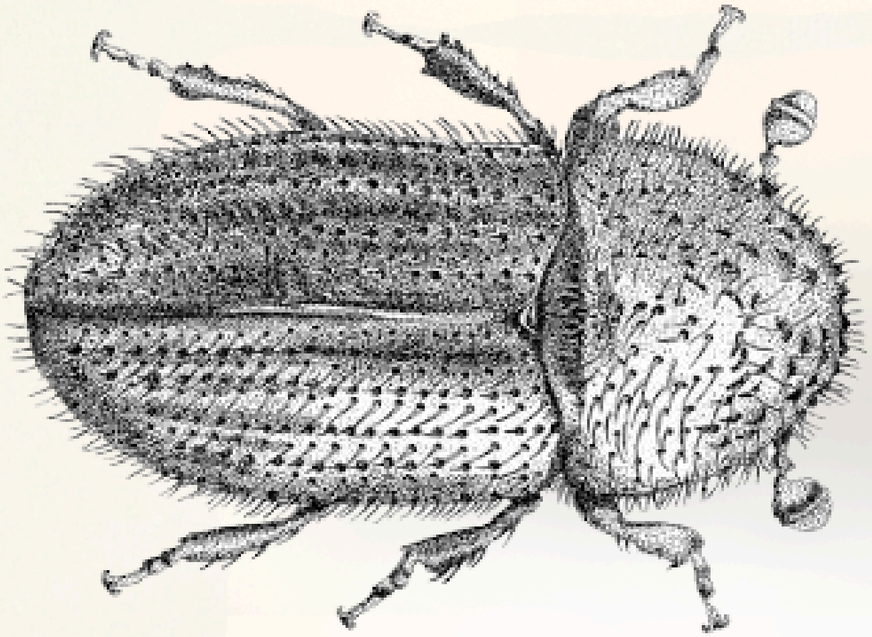
Ciclo de vida: Pasa el 80% dentro del fruto.

Asimetría reproductiva: La generación de hembras es 10 veces mayor a la de machos.

Invasión aérea: Solo las hembras tienen capacidad de vuelo para colonizar frutos

Una plaga originaria de África casi imposible de controlar una vez coloniza el fruto. Necesitamos conocer su genoma para detenerla.

Descifrando la biología del invasor



18.765
Genes descifrados

Navarro-Escalante, L., Hernandez-Hernandez, E.M., Nuñez, J. et al. A coffee berry borer (Hypothenemus hampei) genome assembly reveals a reduced chemosensory receptor gene repertoire and male-specific genome sequences. Sci Rep 11, 4900 (2021). https://doi.org/10.1038/s41598-021-84068-1

Trabajo conjunto entre Cenicafé y la Universidad de Purdue (EE.UU.)

Publicado en Scientific Reports (Nature)

Hito mundial contra la plaga que afecta a 70 países y 20 millones de familias

Resultados



El descubrimiento

191 Genes exclusivos para la percepción de olores y estimulos gustativos.



La vulnerabilidad

Estos genes se activan fuertemente en su estado adulto volador.



Futuras estrategias de control

Diseño de olores falsos que confunden al insecto sin recurrir a plaguicidas.



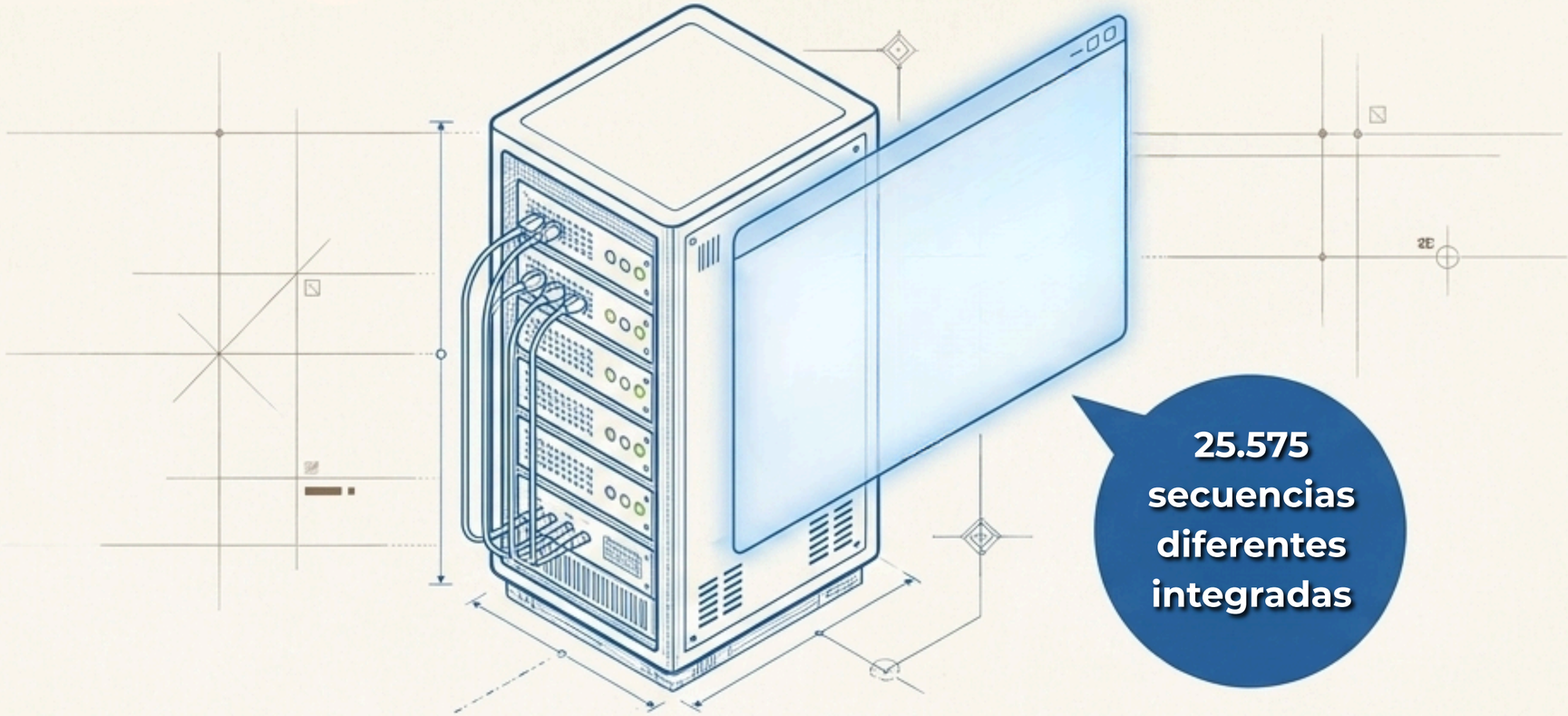
Proyecto FONTAGRO:
USD 1.77 millones (2008)

Impacto: 36 meses de ejecución y fortalecimiento de la red científica.

La tecnología no funciona sola. El verdadero motor de la genómica es humano.

Este esfuerzo capacitó a 4 científicos colombianos incluyendo 4 estudiantes de doctorado en la élite mundial para interpretar datos.

¿Qué nos dejó el proyecto?



25.575 secuencias diferentes integradas

Datos genómicos → Ensamblajes → Bases de Datos Públicas (NCBI) → Patrimonio científico global

Panorama actual

Phymastichus coffea



Fuente: Disciplina de Entomología

Prorops nasuta (hembra)



Fuente: Disciplina de Entomología

Avispas parasitoides utilizadas para el control biológico de la broca del café

Trabajo conjunto entre Cenicafé y USDA

Genomas publicados en el NCBI

Descubrir la baja variabilidad genetica de las colonias en Cenicafé

¿Cuánto vale un genoma?

¿Qué pasa con ese genoma 10 años después?

La acumulación silenciosa

Pregunta biológica

Secuenciador



683.718 archivos

SABEMOS
LO QUE HACEMOS

Cenicafé
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia





¿Dónde está almacenado?

¿Quién puede encontrarlo?

¿Existen copias?

¿Tenemos los datos originales?

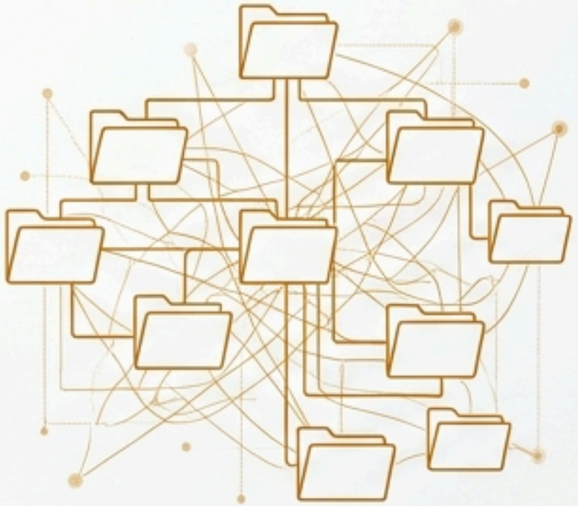
¿Podemos volver a analizarlo?

¿Podemos utilizarlo para responder una pregunta que todavía no existía cuando fue generado?

PARTE IV

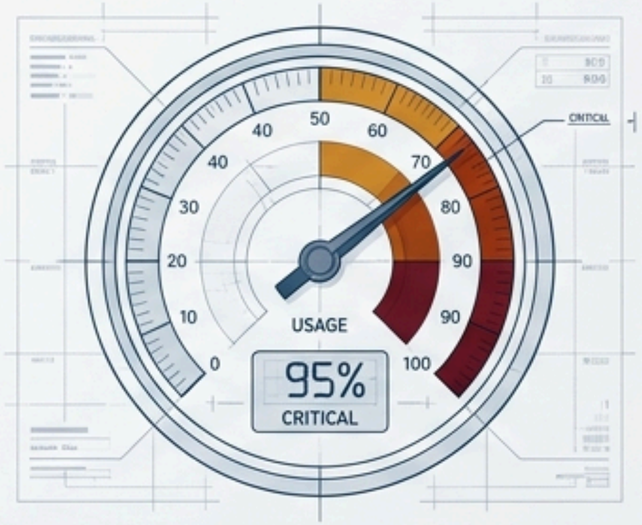
GESTIÓN DEL CONOCIMIENTO

Evolución de la infraestructura NAS



El origen

Durante años, los proyectos generaron grandes volúmenes de datos: Secuenciación, archivos intermedios, ensamblajes y códigos



La realidad

El crecimiento progresivo transformó el almacenamiento en un repositorio histórico pasivo. La información está disponible, pero carece de clasificación funcional o de reutilización.

Los tres riesgos

1. Duplicación estructural y consumo crítico de capacidad.
2. Imposibilidad de localizar rápidamente evidencia cientifica.
3. Conservación indiscriminada de archivos regenerables.

El propósito del inventario

1. 

Inventario

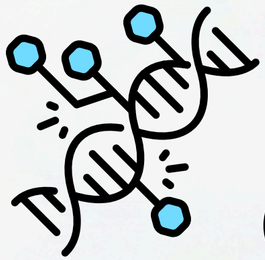
¿Qué información existe?
(Conteo y volumen total)

2. 

Topografía

¿Dónde está ubicada?
(Rutas y ecosistemas)

El objetivo no es revisar carpetas, sino **construir** una **visión cuantitativa** del patrimonio digital para tomar decisiones informadas.

3. 

Taxonomía

¿Qué tipo de información representa?
(Clasificación bioinformática)

4. 

Gobernanza

¿Qué debe conservarse, revisarse o depurarse?
(Política de gestión)

Fase 1: Extracción y estructuración de metadatos



Extracción de datos

Paso 1

Evitar la exploración manual invasiva.

Paso 2

Ejecutar un escaneo sistemático del servidor.

Paso 3

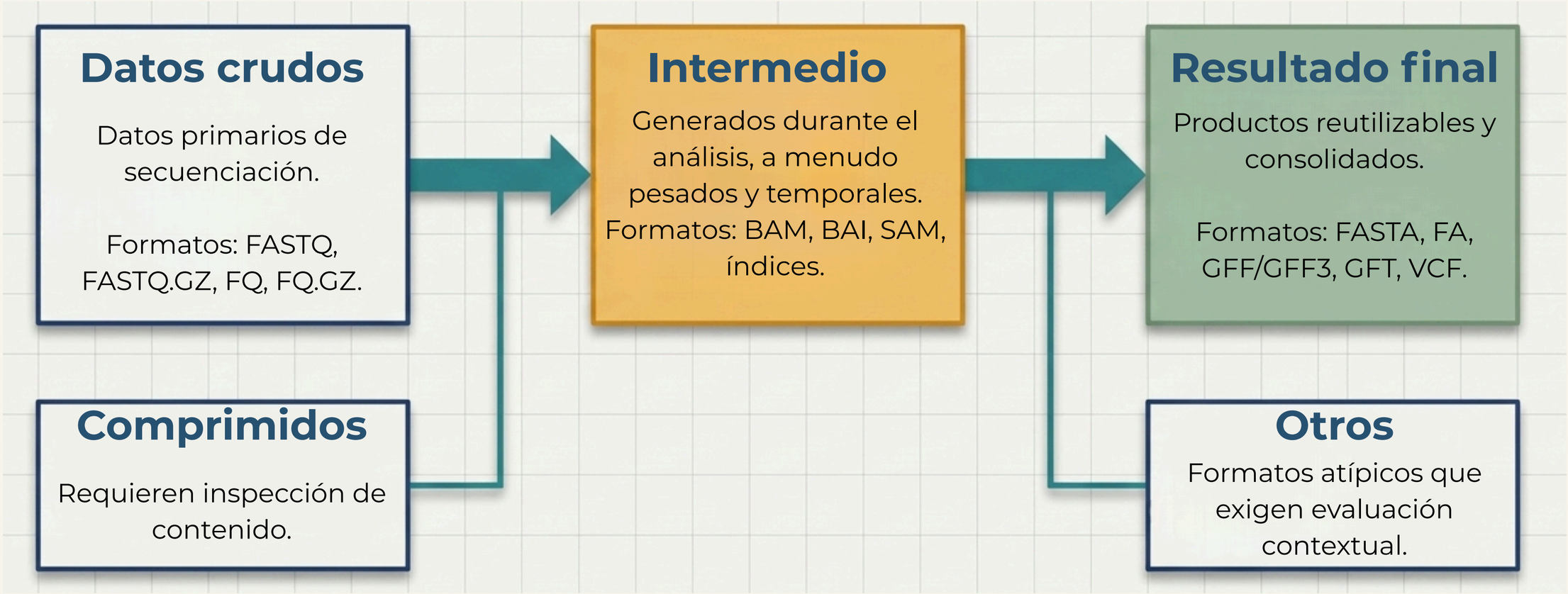
Transformar la jerarquía en un excel estructurado

Path	Filename	Extension	Size	Modified Dete	Structoral Locstion

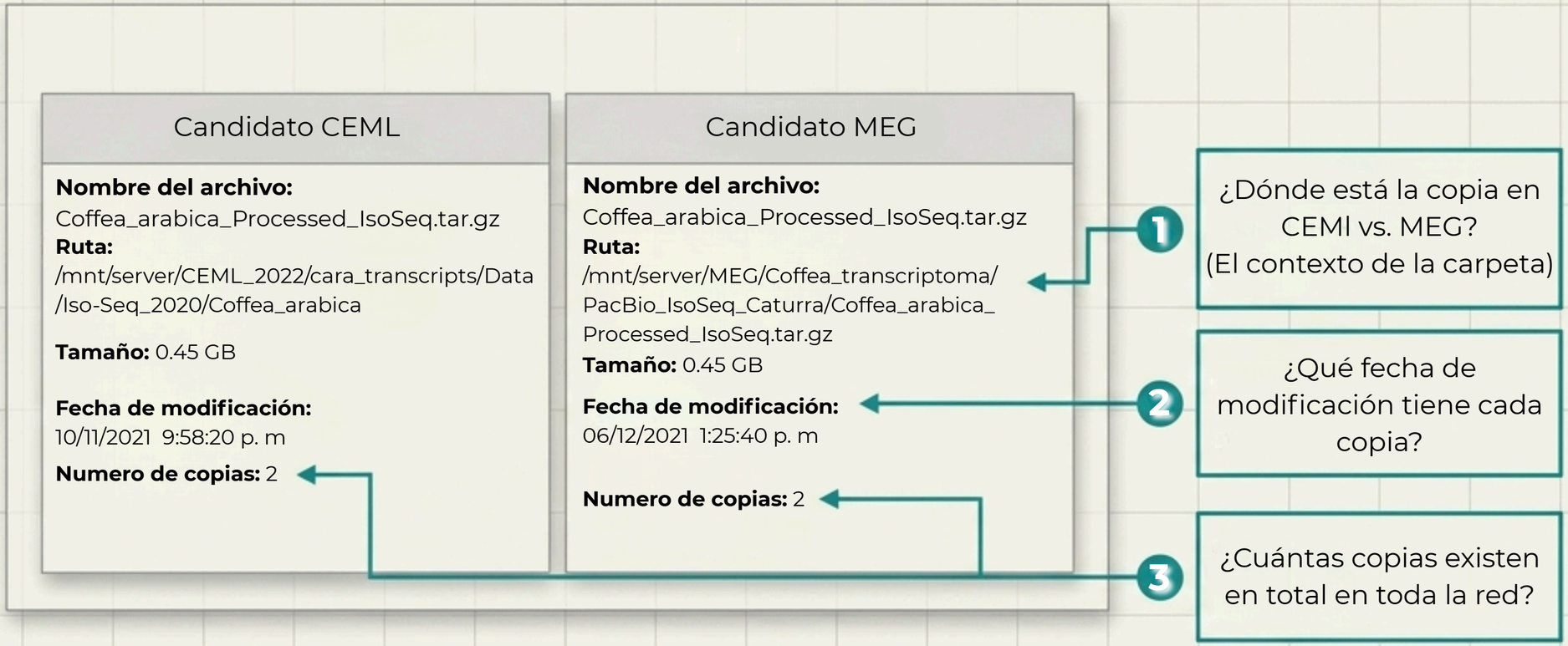
Cada registro capturó variables exactas:

- **Ruta completa (Path)**
- **Nombre de archivo**
- **Extensión**
- **Tamaño**
- **Fecha de modificación**
- **Ubicación estructural**

Fase 2: Clasificación bioinformática



La anatomía de una decisión: Comparación a la par

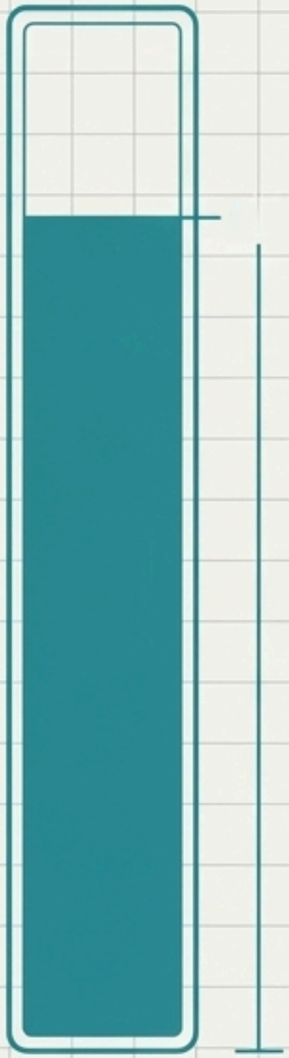


El objetivo no es la eliminación automática. Es identificar cuál es la copia principal y cuál es la redundancia candidata a revisión.

Política de conservación y depuración

Categoría de acción	Definición y criterio	Tipo de archivo objetivo
Conservar	Datos crudos y resultados finales de alto valor. Protección incondicional.	RAW, FINAL RESULT
Revisar	Archivos con dependencias o contexto ambiguo. Requieren validación del investigador.	COMPRESSED, ARCHIVOS COMPLEJOS
Revisar para borrar	Candidatos redundantes o intermedios prescindibles. Prioridad de evaluación.	DUPLICADOS, TEMPORALES
Alta prioridad para revisar	Intermedios antiguos y masivos. Su depuración libera espacio crítico.	INTERMEDIATE (Masivos)
Revisión manual	Formatos atípicos o no clasificados automáticamente.	OTHER

Estado institucional y próximos pasos



683.718

Archivos

Más del 80% de la revisión completada. Se han identificado de forma segura estos archivos para conservación definitiva dentro de la NAS.

Validación criptográfica

Transición de filtros por ‘nombre/tamaño’ a comparación estricta mediante Hashes (SHA-256).

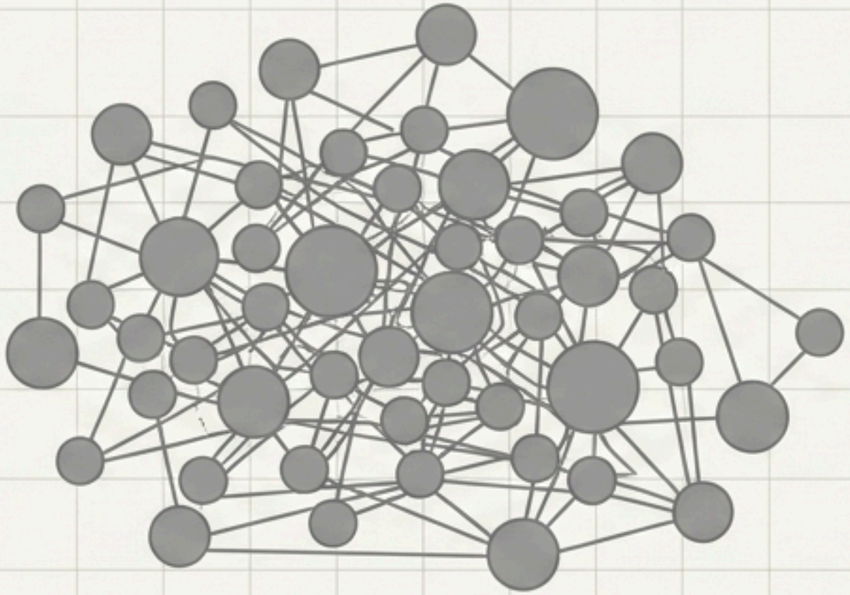
Red de seguridad

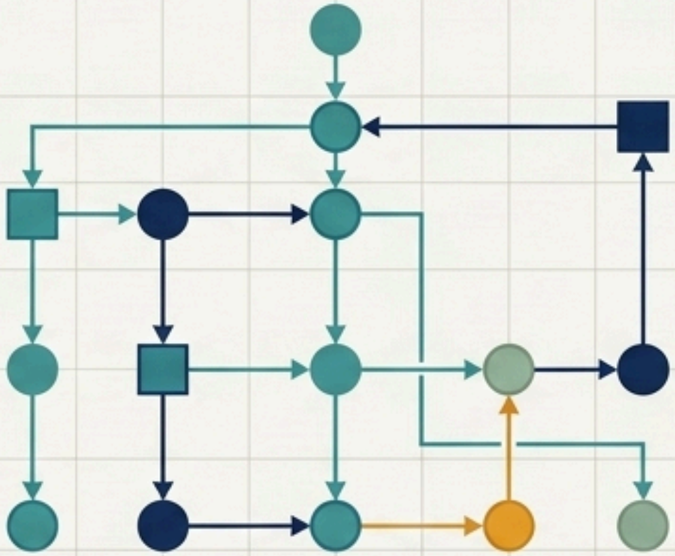
Implemetación de políticas de respaldo institucional previas a cualquier eliminación de datos.

Trazabilidad científica

La eliminación es un proceso controlado y validado por responsables de los datos, no un borrado ciego del sistema.

Un patrimonio gestionado





La depuración es una consecuencia de este conocimiento, no el objetivo principal. Una NAS científica bien gestionada no es solamente un lugar donde se guardan archivos, es parte del patrimonio científico de la institución.

Localización inmediata

Reducción de la redundancia

Reutilizacion de resultados históricos garantizada

PARTE V

CASOS DE ESTUDIO



SABEMOS
LO QUE HACEMOS

Cenicafé
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia

La infraestructura cobra valor únicamente a través de las preguntas biológicas que perite responder. Estas investigaciones operan de manera independiente, pero comparten un mismo motor analítico.

Capacidad computacional



Caso 1

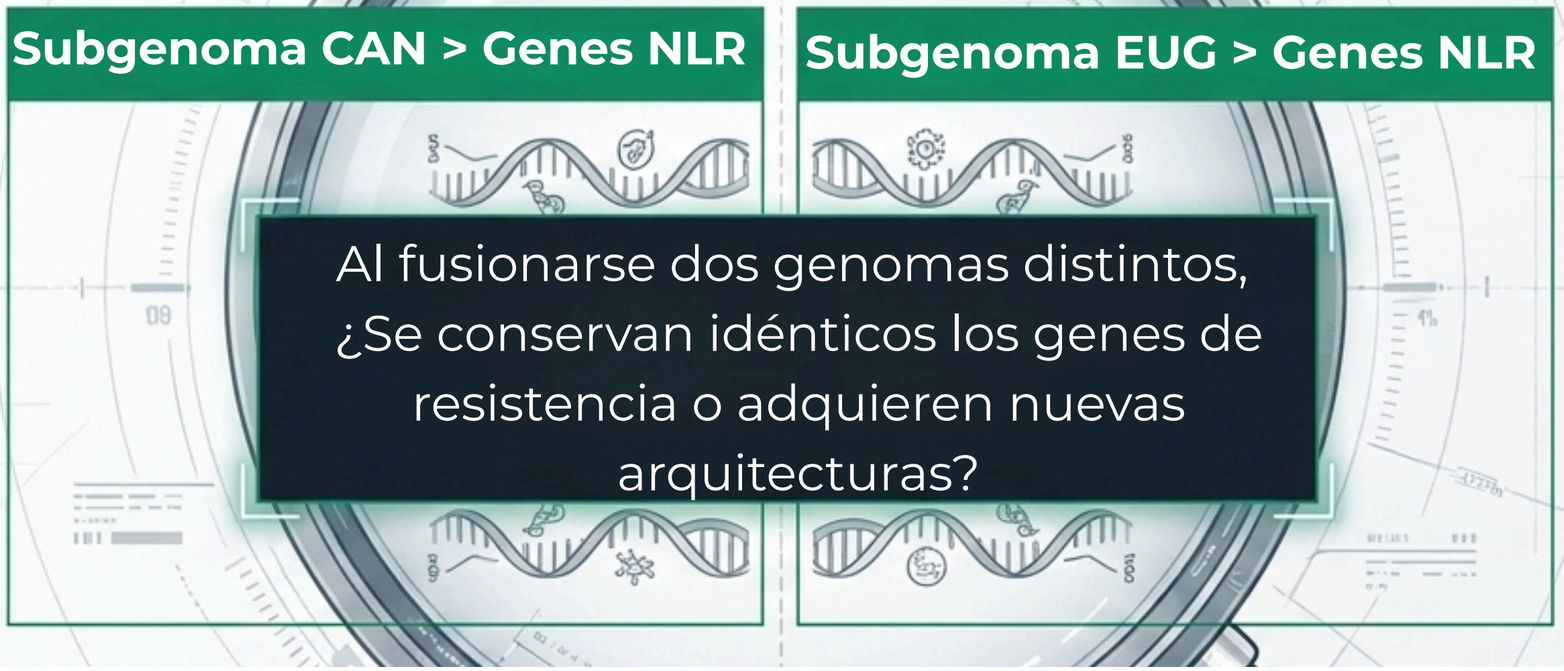
Genómica del café. Dinámicas intra-genoma.



Caso 2

Genómica de la roya. Dinámicas inter-genoma.

¿Qué ocurre con la resistencia tras la fusión genómica?

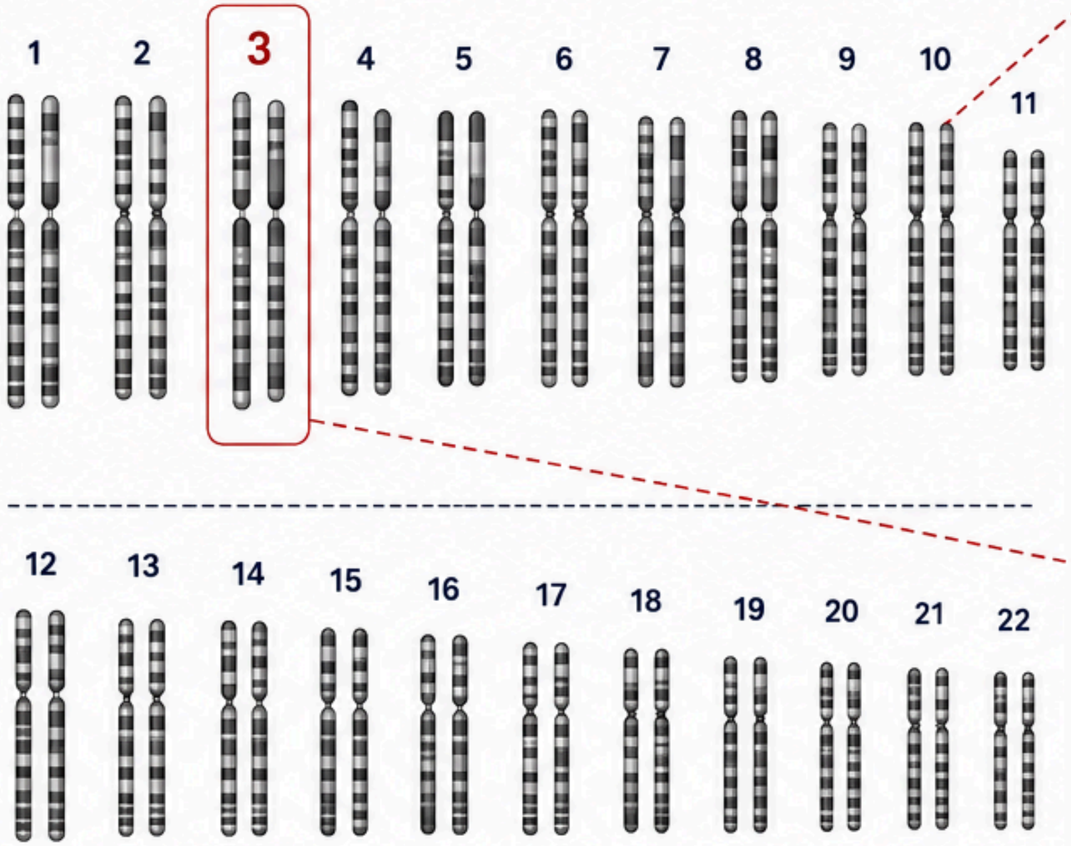


Al fusionarse dos genomas distintos, ¿Se conservan idénticos los genes de resistencia o adquieren nuevas arquitecturas?

Para responder, la bioinformática debe buscar, filtrar y comparar proteínas a escala genómica.

Genoma de *Coffea arabica*

Los 22 cromosomas



Coffea arabica **es un alopoliploide tetraploide**

El genoma está compuesto por 22 pares de crommosomas (11 heredados de *C. canehopra* y 11 de *C. eugenioides*

Nuestro estudio identificó un candidato en el cromosoma 3

Cromosoma 3



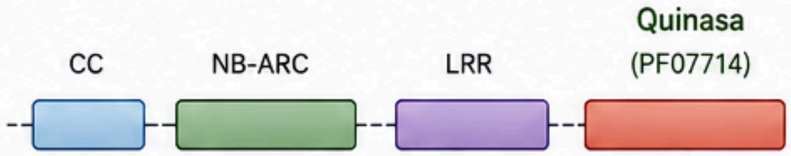
Nuevo indicio en el cromosoma 3

Identificamos un par homeólogo

chr3_can.g16927

(derivado de *C. canephora*)

1,609 aa



Quinasa = interruptor de alarma
Recibe la señal y la transmite, activando la respuesta de defensa de la planta.

chr3_eug.g18075

(derivado de *C. eugenioides*)

1,151 aa



Evidencia de expresión

Las lecturas de RNA confirman que el candidato no está solo escrito en el genoma, sino que está siendo utilizado por la planta.

Evolutionary divergence of homeologous NLRs produced a kinase-integrated immune receptor in Coffea arabica

Authors:

Nicolás Ocampo Moreno[1], Lucio Navarro Escalante[1], Marco Aurelio Cristancho Ardila[1]

Affiliations:

[1] National Center for Coffee Research – Cenicafé, Manizales, 170009, Colombia

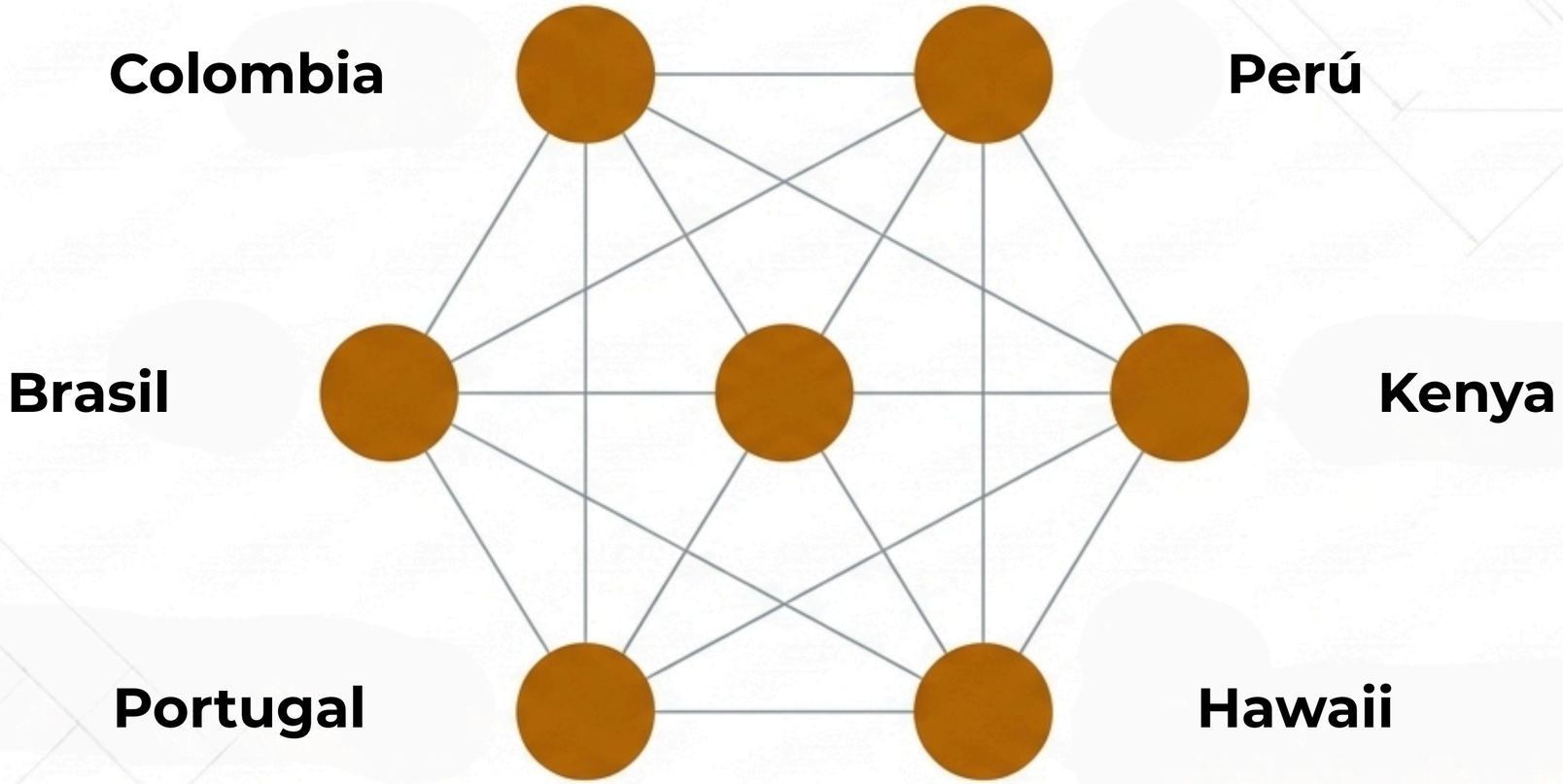
Publicación en desarrollo para someter en PLOS One

Cuando analizar un solo genoma no es suficiente

Transición de la búsqueda intra-genoma (Café) a la genómica comparativa inter-genoma (Roya)

El café se defiende, pero el patógeno también evoluciona

El problema de la Roya: 6 Genomas, 1 pregunta

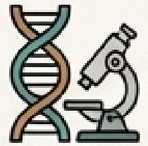


Objetivo: Entender que diferencias existen entre sus poblaciones y qué proteínas utilizan para infectar y superar la resistencia del café.

(Desarmando) a la Roya *(Hemileia vastatrix)*

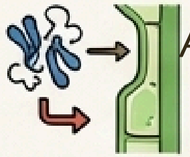


Mapeando al enemigo



Secuenciación y análisis de múltiples aislados y genomas del hongo.

Identificación de efectores



Anotación de proteinas y genes de virulencia, herramientas el hongo para infectar.

Rastreando la evolución



Comprendiendo como muta el hongo para desarrollar nuevas aplicaciones preventivas,

Comparative genomics of geographically diverse *Hemileia vastatrix* isolates provides candidate effectors for avirulence gene discovery

Nicolas Ocampo-Moreno1,+, Simon Orozco-Arias2,3,+,, Simon Gaviria-Orrego3, Isabel Echeverri-Ocampo4, Lucio Navarro-Escalante1,Savina Gutierrez6, Eveline Caixeta7, Dora Batista8, Alfredo Herrera9, Romain Guyot4,5, Marco Cristancho1,*

1 National Center for Coffee Research - Cenicafé, Manizales, 170009, Colombia
2 Life Sciences Department, Barcelona Supercomputing Center, Barcelona, 08034, Spain
3 Department of Computer Science, Universidad Autónoma de Manizales, 170001, Manizales, Colombia
4 Department of Electronics and Automation, Universidad Autónoma de Manizales, 170001, Manizales, Colombia
5 Institut de Recherche pour le Développement (IRD), UMR DIADE, Université de Montpellier, CIRAD, Montpellier, France

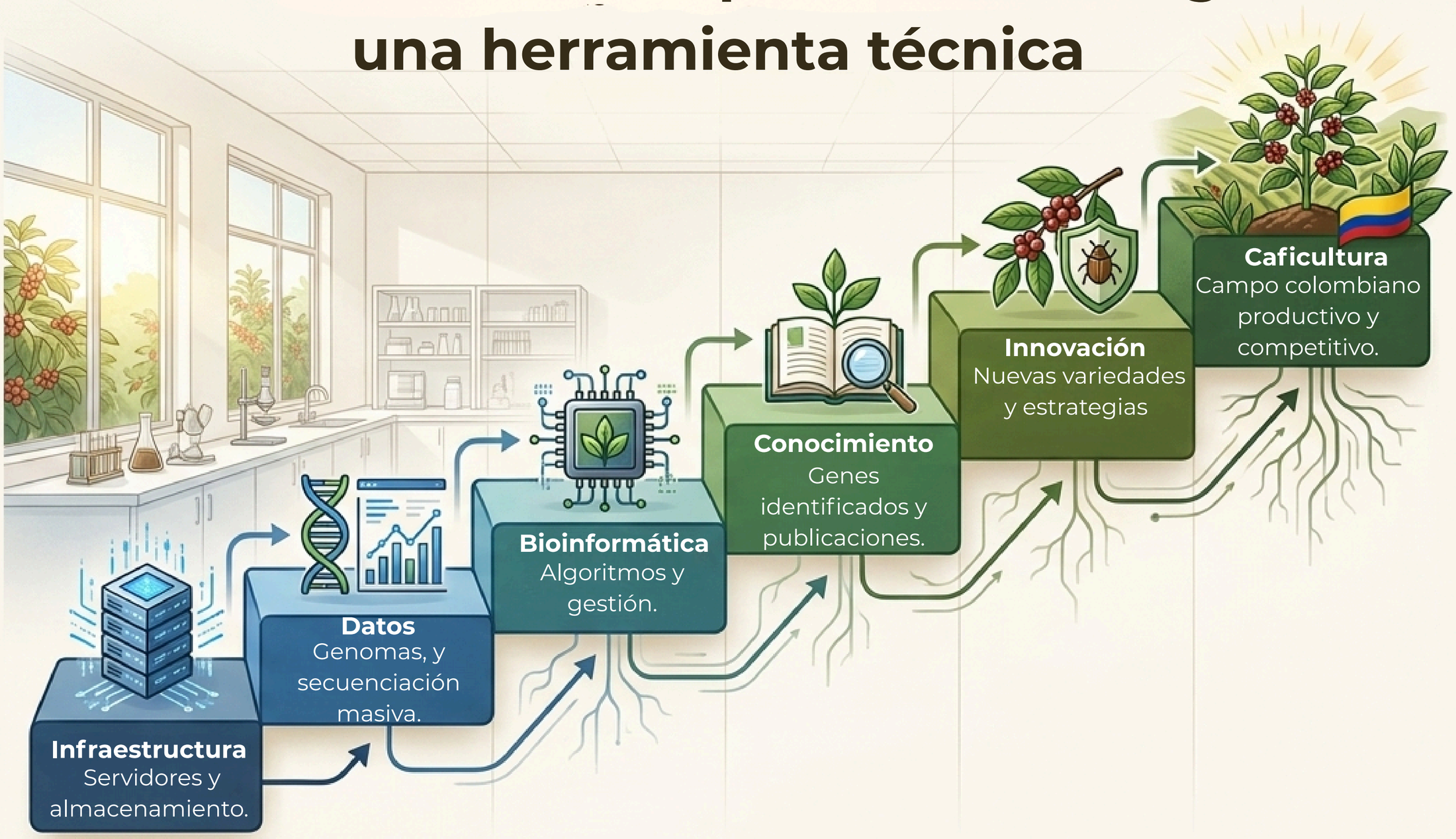
Publicación en desarrollo para someter en BMC Genomics



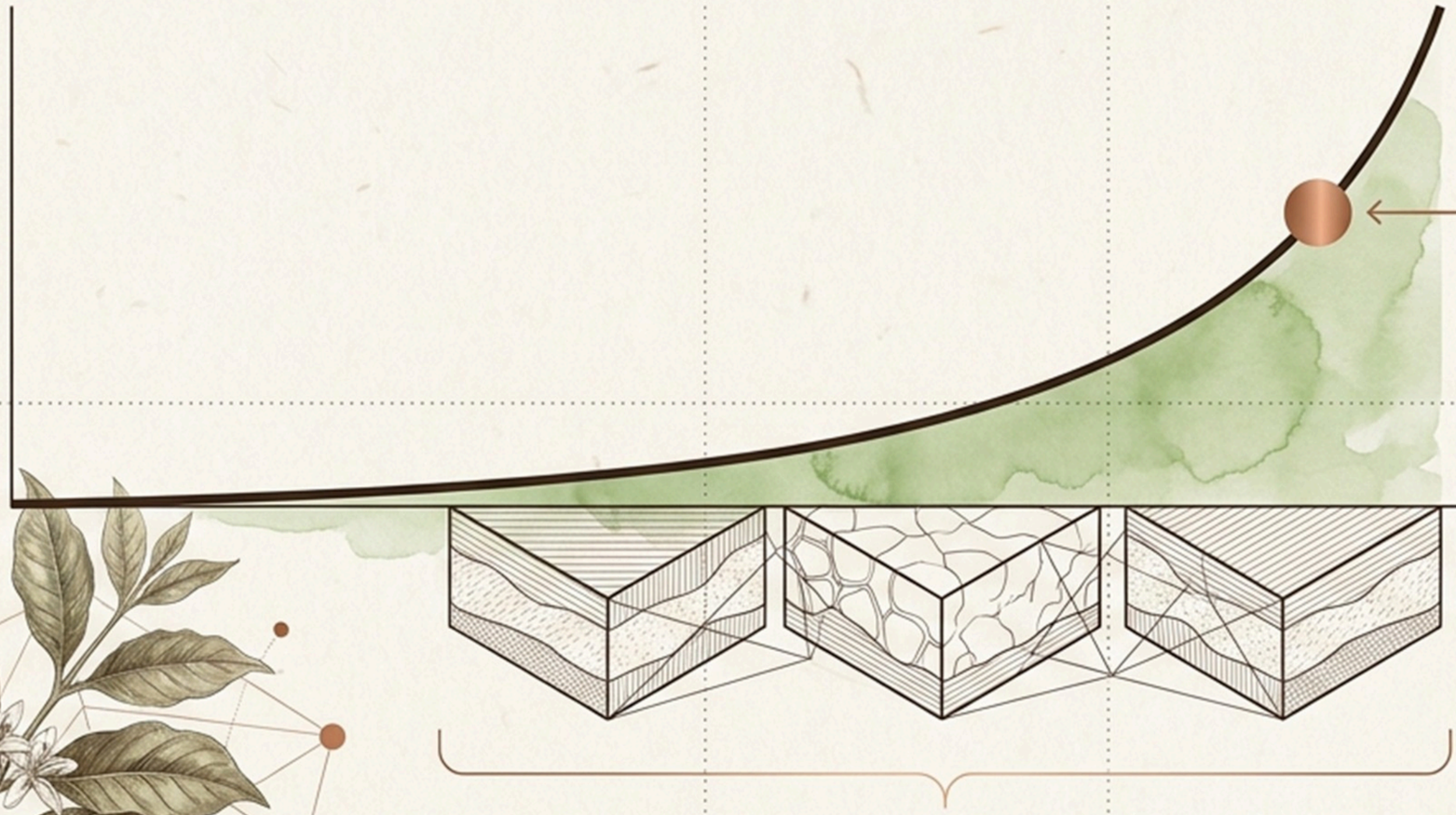
PARTE VI

EL FUTURO EN CENICAFÉ

La bioinformática es una capacidad estratégica no solo una herramienta técnica



La consolidación de nuestra capacidad analítica



Presente

Esfuerzo sostenido

La inversión realizada durante estos años en los analisis bioinformaticos.

Una base robusta, construida a traves del esfuerzo sostenido que hoy da soporte a la investigación

El destino estratégico



Colocar a Cenicafé en el estándar internacional de la genómica del café

Agradecimientos



SABEMOS
LO QUE HACEMOS

Cenicafé
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia







SABEMOS
LO QUE HACEMOS

Cenicafé
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia



SABEMOS
LO QUE HACEMOS

Cenicafé
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia






BORN REBEL



Cenicafé®
Centro Nacional de Investigaciones de Café

Federación Nacional de Cafeteros de Colombia®

cenicafe.org Cenicafé FNC @cenicafe cenicafé CenicaféFNC @cenicafefnc