



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

EL PERÚ PRIMERO

***“Análisis de Polimorfismos en el Genoma de Cuy  
(Cavia porcellus) asociados a  
características de Importancia Económica”  
Proyecto 171\_PI***

**Blga. Fredesvinda Carrillo Castillo**  
Investigadora Responsable del Proyecto



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

# “Análisis de Polimorfismos en el Genoma de Cuy (*Cavia porcellus*) asociados a características de Importancia Económica”

Periodo de Ejecución 2018-2020

**Importe total del Proyecto : S/ 666,450**  
**(Seiscientos sesenta y seis mil cuatrocientos cincuenta con 00/ soles)**  
**Aporte PNIA: S/ 500,000**

**PNIA**  
Programa Nacional de  
Innovación Agraria





PERÚ

Ministerio  
de Agricultura y Riego

## Antecedentes

- El cuy (*Cavia porcellus*) representa un valor cultural, religioso y económico de la población peruana y últimamente el consumo de su carne se ha extendido a diferentes regiones de Norteamérica, Asia y África.
- La carne de cuy presenta buenas características nutritivas: 19.1 % de proteínas y 7.41 % de grasa.
- En el 2016, Perú exportó más de 15 mil toneladas de carne a Norteamérica y Asia.





PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Antecedentes

- Los primeros estudios en cuyes se realizaron utilizando marcadores mitocondriales y microsatélites, resultados del proyecto O91\_PI: “Utilización de herramientas moleculares para la caracterización genética de las razas de cuy Perú, Inti y Andina generadas por el INIA”.
- La Diversidad genética encontrada en cuyes nativos fue más alta que en las razas mejoradas del INIA.
- Las razas Perú, Inti y Andina se diferenciaron genéticamente utilizando 30 marcadores microsatélites.
- Actualmente, el secuenciamiento del genoma es una de las estrategias para la detección de variantes que permiten identificar regiones de interés en la expresión de características productivas.

***Colección de ADN de cuyes nativos de 6 departamentos: Junín, Huancavelica Apurímac, Cusco, Puno, Cajamarca.***





PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Antecedentes

- A diferencia de otras especies animales bajo selección, para el caso del cuy, no se cuenta con información de los genes o marcadores asociados a características productivas, lo que permitiría incrementar la precisión de selección en los programas de mejoramiento genético.
- El proyecto “Análisis de polimorfismos en el genoma de cuy (*Cavia porcellus*) asociados a características de importancia económica”, tiene como propósito identificar regiones a lo largo del genoma de cuy asociadas a características de importancia económica, mediante tecnologías de Secuenciamiento de Nueva Generación (NGS) y análisis bioinformáticos.





PERÚ

Ministerio  
de Agricultura y Riego

## Objetivos

**O.1 Toma de muestra y validación de  
caracterización de productividad**

**O.2 Análisis de composición genética y  
estructura poblacional del cuy**



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## O.1 Toma de muestra y validación de caracterización de productividad

**O.1.1. Toma de muestra de material biológico de cuyes:** muestras de folículos pilosos colectadas del Galpón de cuyes de la EEA Chumbibamba, Andahuaylas, Apurímac.





PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## O.1 Toma de muestra y validación de caracterización de productividad

### O.1.2. Validación de datos de caracterización de productividad.

**Equipo Técnico  
Responsable de la  
instalación y  
mantenimiento del  
Galpón de cuyes  
nativos**

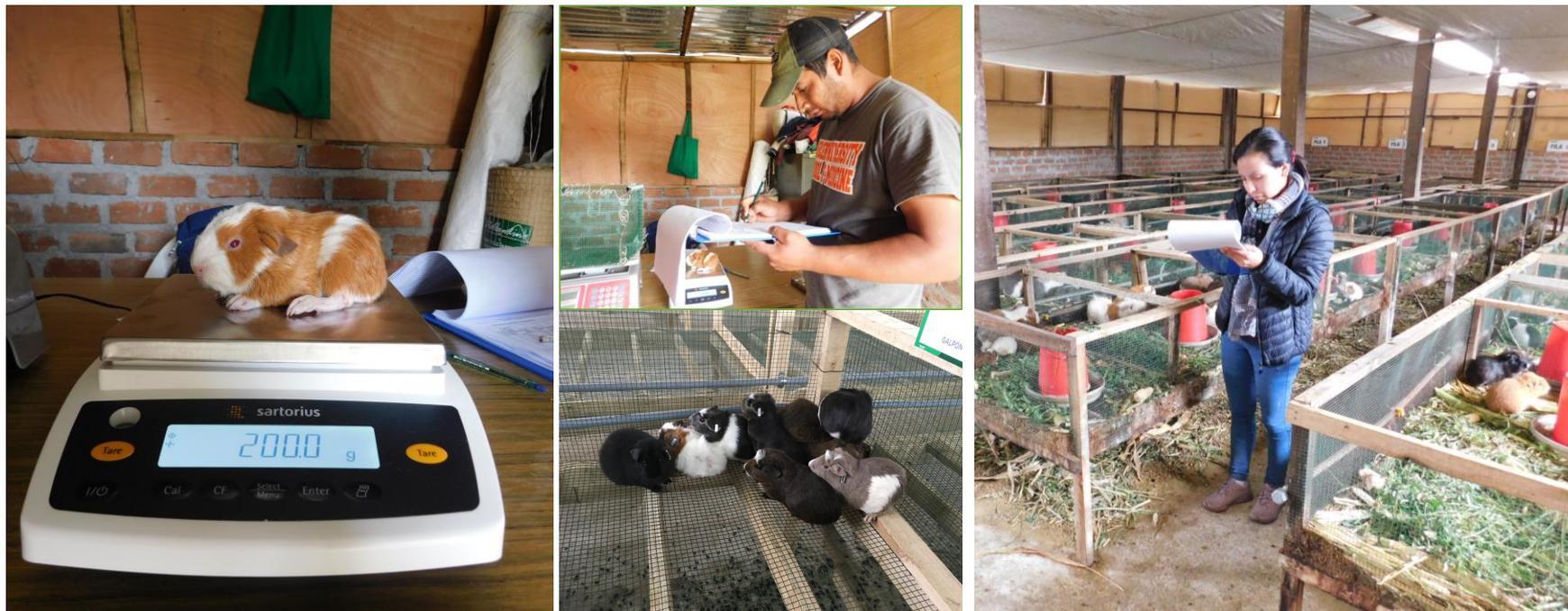
- Ing. Juan Huayhua,
- Darwin Huamán,
- Roy Sarmiento



**Instalación de un Galpón en la EEA Chumbibamba de cuyes nativos de Apurímac**



## O.1.2. Validación de datos de caracterización de productividad.



**Recolección de datos productivos de cuyes nativos - Galpón en la EEA Chumbibamba, Andahuaylas, Apurímac**



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## O.2 Análisis de composición genética y estructura poblacional del cuy

### O.2.1 Extracción de ADN y preparación de muestras para secuenciamiento

Se compararon metodología convencional y comerciales para la evaluación de calidad de ADN extraídas de diferentes tipos de tejidos.

### O.2.2 Secuenciamiento y ensamblaje del genoma del cuy Raza Perú

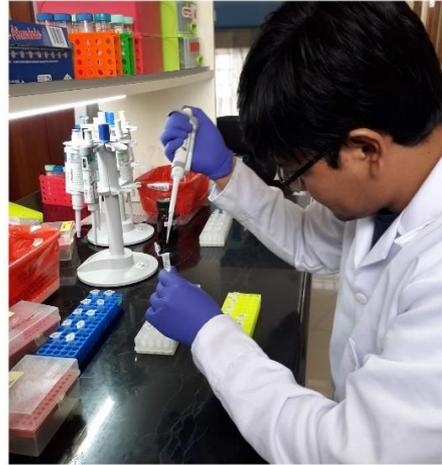
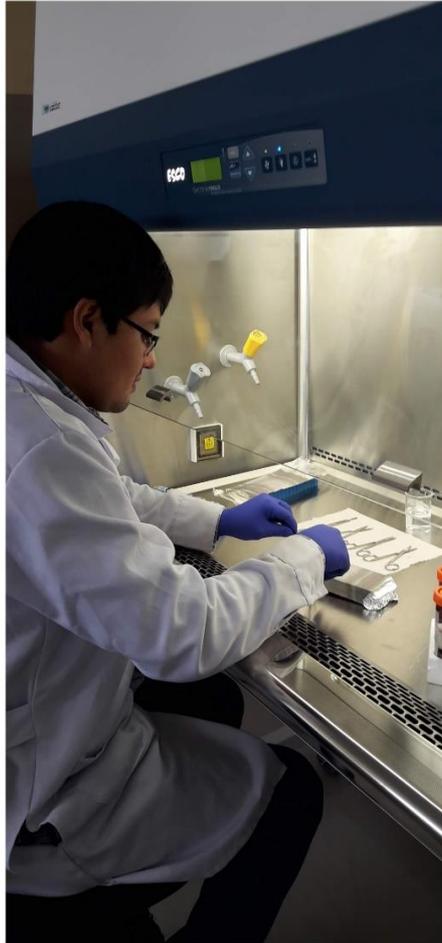
Se secuenciaron usando la plataforma illumina HiSeq 4000, el secuenciamiento se realizó a 10 individuos: 4 razas Perú, 3 silvestres y 3 nativos de Apurímac.





PERÚ

Ministerio  
de Agricultura y Riego



## ***Extracción de ADN y preparación de muestras para secuenciamiento***

## Extracción de ADN

- Para la extracción de ADN se evaluó el protocolo convencional y protocolos comerciales.
- Extracción de ADN mediante Purelink™ Genomic mini kit.
- Se realizó la extracción de muestras de músculo, folículo piloso y muesca de oreja (cartílago).

### Valores de medición de la cantidad y calidad de ADN extraídas de 10 muestras de cuyes

Código	Tejido	Procedencia	Concentración			
			(ng/μl) NanoDrop	260/280	260/230	Concentración (ng/μl) Qubit
CP-1135	Músculo	EEA La Molina	59.24	1,84	2,15	58.4
CP-1136	Músculo	EEA La Molina	79.63	1,89	2,31	67
CP-1137	Músculo	EEA La Molina	32.00	1,9	2,02	30.1
CP-1144	Músculo	EEA Chumbibamba	98.57	1,87	2,19	98.6
CP-1145	Músculo	EEA Chumbibamba	59.25	1,84	2,42	60
CP-1146	Músculo	EEA Chumbibamba	137.1	1,9	2,37	132
CP-1068	Cartílago	EEA Chumbibamba	314.6	1.91	1.9	120
CT-034	Folículo Piloso	Pachachaca- Abancay Apurímac	30,09	1,84	2,07	24.80
CT-035	Músculo	Refugio de vida Silvestre Pantanos de Villa	37.81	1,88	2,27	33,9
CT-036	Músculo	Refugio de vida Silvestre Pantanos de Villa	69.16	1,86	2,35	56,2

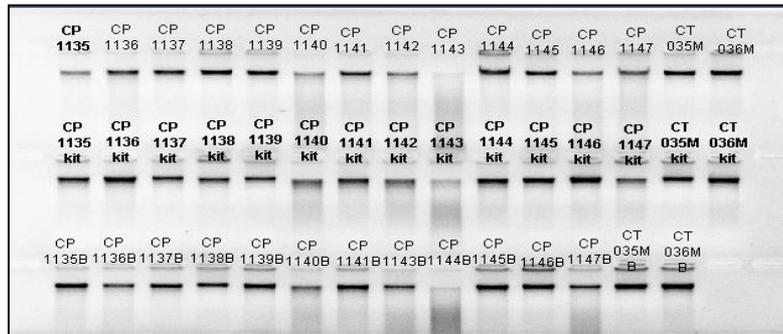
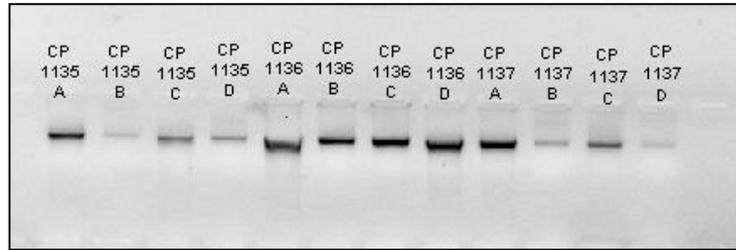
Se comparó las mediciones de concentración de ADN, obtenidas con:

- Espectrofotometría: NanoDrop™ 8000 (Thermo Scientific).
- Fluorometría: Qubit® 3.0 (Thermo Scientific™).



## Evaluación de la Calidad de ADN en geles de agarosa al 1.2%.

Se seleccionaron los ADN de buena calidad, cantidad y sin degradación (Ejemplo: CP 1135).





PERÚ

Ministerio  
de Agricultura y Riego

***Secuenciamiento  
y  
ensamblaje del  
genoma  
del cuy Raza Perú***

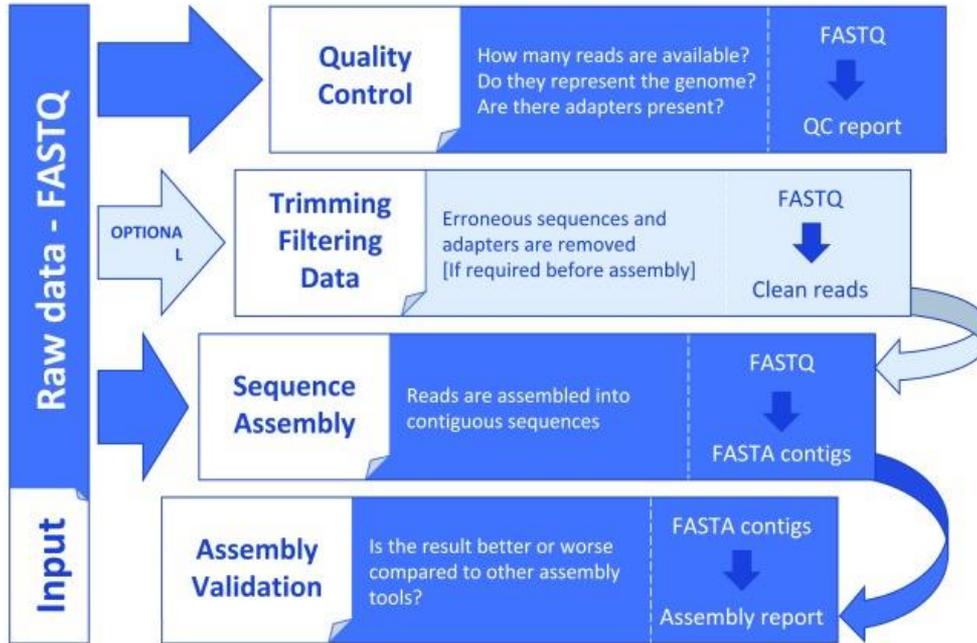


PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria



## Flujograma para el análisis del secuenciamiento de genoma completo de cuyes.

- Evaluación del control de calidad con el programa FastQ (Q30).
- Filtrado *Trimming* para seleccionar aquellas secuencias que tengan un mínimo de calidad igual o superior a Q30.
- Ensamblaje (unión, empalme) de secuencias de buena calidad .

Dominguez Del Angel, V., Hjerde, E., Sterck, L., Capella-Gutierrez, S., Notredame, C., Vinnere Pettersson, O., Amselem, J., et al. (2018). Ten steps to get started in Genome Assembly and Annotation. *F1000RESEARCH*, 7.



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Resumen de la calidad de secuenciamiento de genoma completo de 10 muestras de cuyes

Código	Bioanalizador tamaño (bp)	ID de muestra (*)	Secuencias Totales (**)	Media % Q30
CP-1135	464	SA42979	135.152.047	93,91
CP-1136	440	SA42980	119.965.953	93,95
CP-1137	482	SA42981	128.122.069	92,83
CP-1144	455	SA42911	117.340.314	93,6
CP-1145	449	SA42912	130.787.116	94,1
CP-1146	448	SA42913	140.368.353	94,22
CP-1068	426	SA42914	138.641.244	93,71
CT-034	446	SA42978	116.493.452	93,62
CT-035	456	SA42976	119.162.747	93,56
CT-036	441	SA42977	136.615.724	93,63

- Medición de calidad de secuenciamiento (Q30) es la posibilidad de error de secuenciamiento de 0.001 .

\*ID de muestra= códigos para el análisis, \*\* Total de numero de insertos generados en el secuenciamiento.



PERÚ

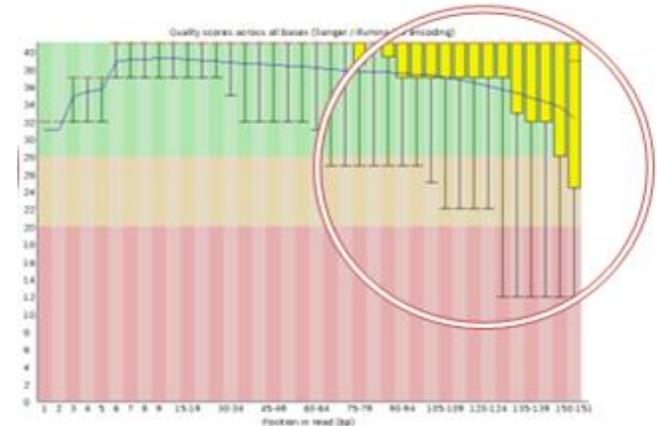
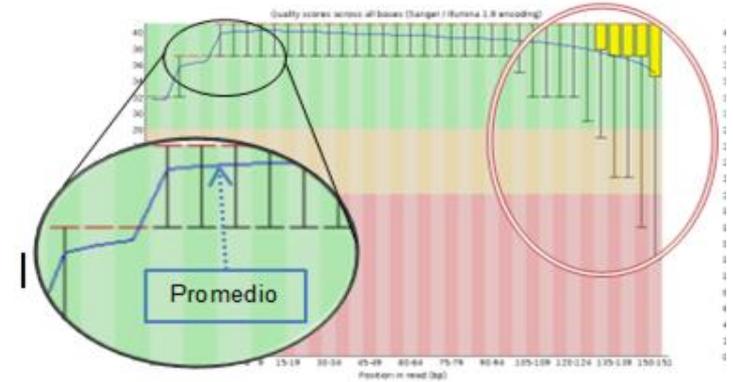
Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Análisis de calidad de secuenciamento, con el programa FastQC

- Se representa en amarillo los cuartiles.
- La línea azul es la mediana o promedio y en rojo, la media de la calidad.
- En el eje X, representa las bases de las lecturas y cada lectura tiene 151 bases.
- Mientras que en el Y, se representan las calidades 0-40, distinguiéndose tres zonas:
  - **Zona verde:** 29-40: Zona de muy buena calidad.
  - **Zona naranja:** Zona de calidad intermedia (20-28).
  - **Zona roja:** Zona de mala calidad (0-20).





PERÚ

Ministerio  
de Agricultura y Riego

# SkimGBS



Instituto Nacional de Innovación Agraria

## *Skim-based Genotyping by Sequencing*

- Determina las diferencias en la composición genética de un individuo.
- Combina dos métodos: genotipado y secuenciamiento de siguiente generación (NGS).

### 1. Identificación de variantes genómicas

- Alineamiento de secuencias al genoma de referencia de *Cavia porcellus* (CavPor3.0 GB:GCA\_000151735.1) utilizando el programa BWA.
- La identificación de SNPs se realizó con los programas SAMtools/BCFtools y GATK.

### 2. Predicción de efectos de variantes genómicas

- Los efectos de las variantes filtradas sobre los genes de *C. porcellus* fueron evaluadas con SNPEff v4.0.
- Se calcularon valores de Ts/Tv y Syn/Non (Transiciones/transversiones, Sinónimas y no sinónimas).



PERÚ

Ministerio  
de Agricultura y Riego

## SkimGBS



Instituto Nacional de Innovación Agraria

- La identificación preliminar de SNPs produjo un total 39 millones de variantes entre los 10 individuos.
- El filtro de variantes con **BCFTOOLS** produjo un set de 759 mil variantes de alta confiabilidad.
- El análisis de los alineamientos con **GATK** produjo 819 mil variantes de alta calidad.
- Al comparar los dos sets de datos, se obtuvo una lista consenso de 746,665 variantes a lo largo del genoma con una densidad promedio de 1 variante cada 3600 pares de bases.
- Ambos métodos de identificación de variantes tuvieron una congruencia del 89.9%.
- Estas variantes incluyen 704,900 SNPs; 24,065 inserciones y 17,700 deleciones.

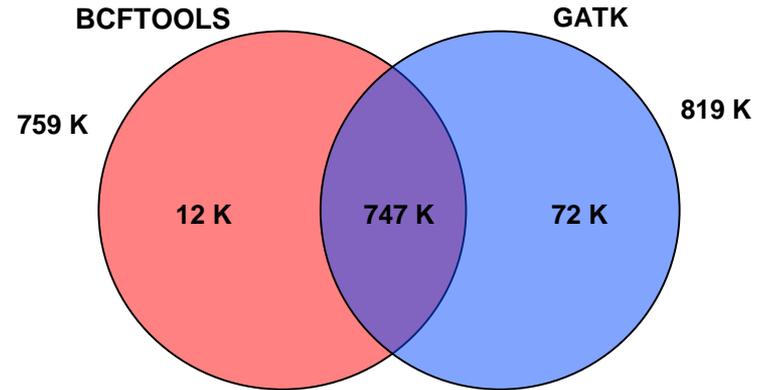


Diagrama de Venn representando el número de variantes obtenidas con diferentes métodos y el número de variantes comunes. El 89.9% de las variantes fueron comunes entre ambos métodos.

**Número total de transiciones, transversiones y ratio Ts/Tv por cada individuo**

Muestra	Transiciones	Transversiones	Ts/Tv
CP-1068	482520	244992	1.970
CP-1135kit	506629	257650	1.966
CP-1136kit	494854	251449	1.968
CP-1137kit	505413	256352	1.972
CP-1144kit	471970	241017	1.958
CP-1145kit	491669	250547	1.962
CP-1147kit	488466	247797	1.971
CT-034kit	467853	239328	1.955
CT-035kit	476982	240739	1.981
CT-036kit	485490	245628	1.977
<b>Total</b>	<b>4871846</b>	<b>2475499</b>	<b>1.968</b>

El *ratio* de transiciones a transversiones (Ts/Tv) para SNP bialélicos fue de 1.968 para las 10 muestras, el cual es el valor típico encontrado en poblaciones de mamíferos.



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria



## Genoma Mitocondrial de *Cavia porcellus*



PERÚ

Ministerio  
de Agricultura y Riego



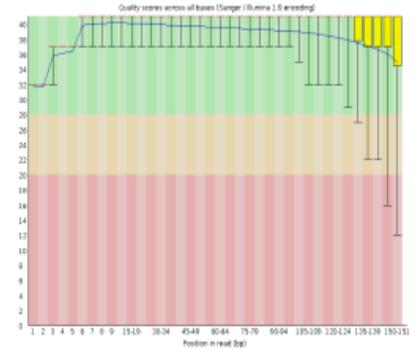
Instituto Nacional de Innovación Agraria

# Genoma Mitocondrial de *Cavia porcellus*

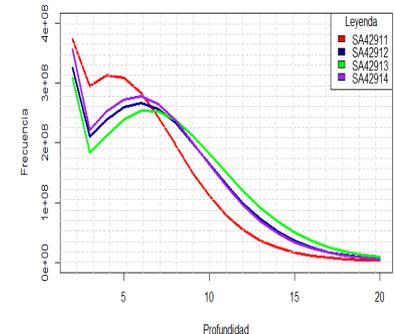
Se analizaron 10 genomas de cuyes: raza Perú (4), nativo de Apurímac (3) y silvestre (3).

1. Pruebas de calidad de secuencias FastQ.
2. Determinación de k meros (subfragmentos de ADN de un tamaño determinado K), con la herramienta Jellyfish.
3. Alineamiento al genoma de referencia (Cavpor v. 3.0 ) Bowtie2
  - Filtrado de Lecturas Mitocondriales con el programa Trimomatic
  - Gráficos de Profundidad con programas SAMtools, Rstudio
4. Ensamblaje de Genoma mitocondrial con el programa SPAdes
5. Validación de ensamblaje y anotación de genoma con el programa QUAST, MITOwebserver y Geseq.

Per base sequence quality

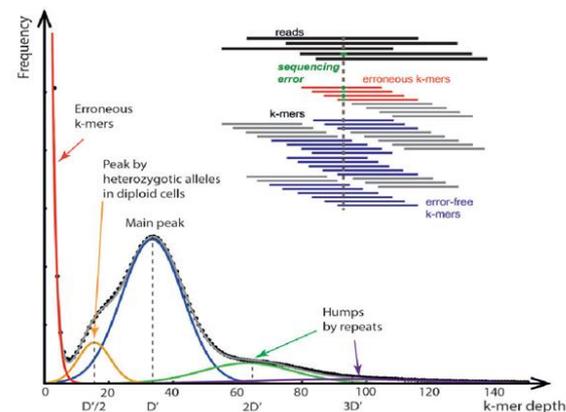
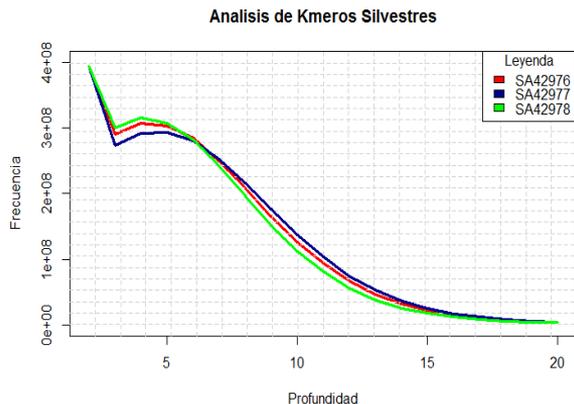
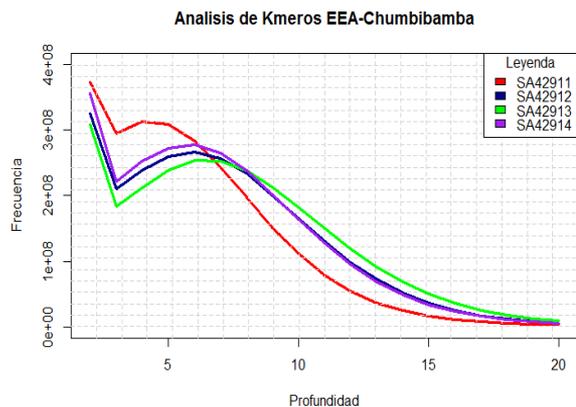


Análisis de Kmeros EEA-Chumbibamba



## Comparación de k-meros entre las muestras raza Perú, nativo de Apurímac y Silvestre.

- Individuos silvestres y el individuo de código SA42911 con los menores picos de profundidad (Número de copias promedio por posición del genoma).
- Las muestras extraídas de tejido muscular produjeron una mayor profundidad de secuenciamiento, seguido por folículo piloso y cartílago.



**K-meros.** Esta secuencia k-mer es una secuencia de un número k de bases que hace parte de la longitud total de la lectura.



PERÚ

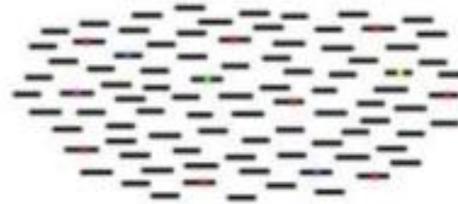
Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

### Profundidad de Cobertura.

Número de lecturas que cubren una base, es decir número de veces que se ha secuenciado una base, es uno de los factores determinantes para evaluar la fiabilidad del nucleótido asignado a esa posición del genoma.

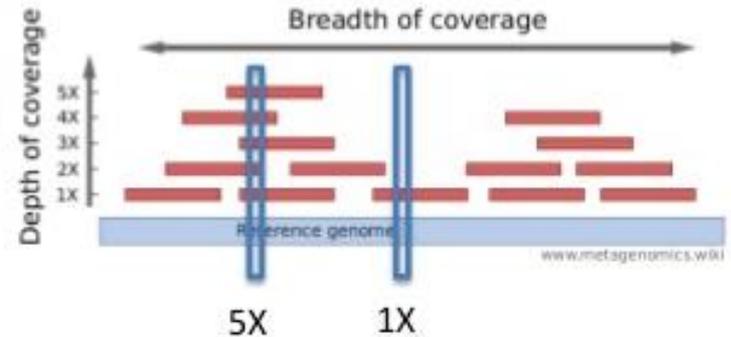


¿que debo de exigir a mi secuencia?

**Profundidad**

20X, 40X 100X

**Cobertura**



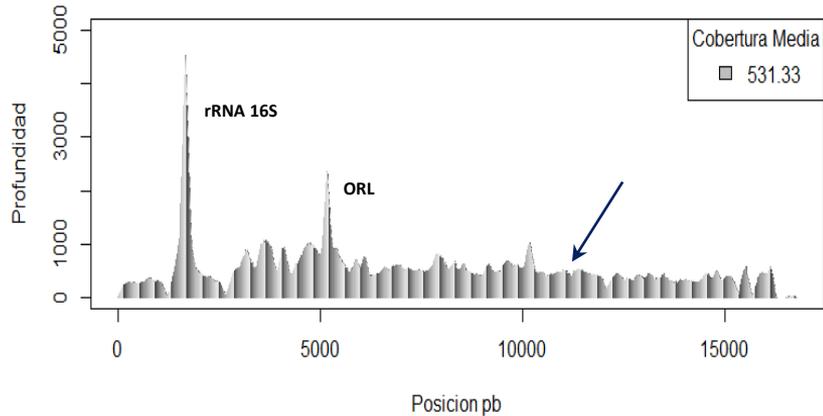


PERÚ

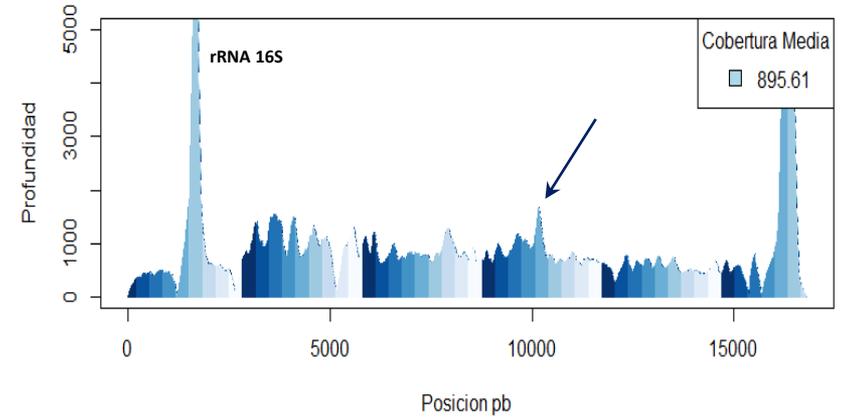
Ministerio  
de Agricultura y Riego

## Profundidad de Cobertura del genoma mitocondrial de *Cavia porcellus*

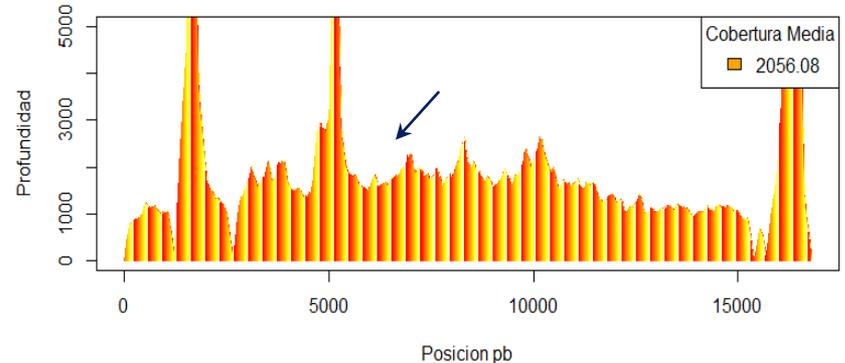
Genoma Mitocondrial SA42912



## Genoma Mitocondrial SA42976



## Genoma Mitocondrial SA42981



### Profundidad de cobertura de genoma mitocondrial de Regiones con mayor profundidad:

- Secuencia del gen de ARN ribosomal 16S.
- Origen de replicación de la cadena ligera (ORL)
- Las flechas indican el gen NADH deshidrogenasa subunidad 4L.



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Ensamblaje de genoma mitocondrial

*“Al proceso de descifrar la secuencia genómica a partir de pequeños fragmentos de ADN, en conjunto con alguna información adicional disponible, se le denomina ensamblaje de genomas.*

Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas bacterianos: una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*, 37(2), 125-132.

Se realizó una reconstrucción del genoma mitocondrial de *Cavia porcellus* (cuy doméstico), para la reconstrucción se utilizaron datos de secuenciamiento Illumina, realizando una comparativa con el genoma de referencia publicada en la base de datos del NCBI (NC\_000884.1) con el programa **SPAdes**.



PERÚ

Ministerio  
de Agricultura y Riego

Instituto Nacional de Innovación Agraria

## Resumen de ensamblaje de genoma mitocondrial

	Código	Tamaño total de ensamblaje (pb)	Numero de <i>Contigs</i>	<i>Contig</i> más Largo (pb)	<i>Contig</i> más Corto (pb)	Total reads (*)
EEA - Chumbibamba	SA42911	16533	1	16533	-	42054
	SA42912	16335	1	16335	-	45322
	SA42913	16432	1	16432	-	31942
	SA42914	16335	2	15696	681	21196
Silvestres	SA42976	16393	1	16393	-	71692
	SA42977	16541	2	15265	1276	17332
	SA42978	16575	2	11915	4660	17094
EE-La Molina	SA42979	16747	1	16747	-	9000(**)
	SA42980	16414	1	16414	-	77710
	SA42981	16402	1	16402	-	20000(**)

- Se obtuvo el ensamblaje de 10 genomas mitocondriales no circularizados.
- Se obtuvieron 2 *contigs* como máximo de ensamblaje. La secuencia de mayor longitud fue de 16747 pb (SA42979) y de menor longitud fue de 681pb (SA42914).
- Se obtuvieron mejores ensamblajes con subconjuntos de lecturas en las muestras SA42979 y SA42981.



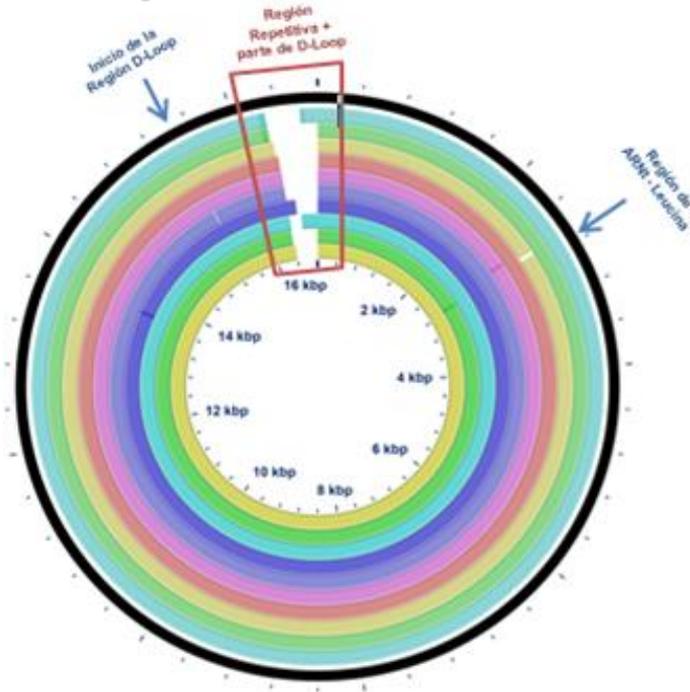
PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Comparativa de las secuencias



Gráfica de BLAST de las secuencias ensambladas con el programa SPAdes al genoma de Referencia de *Cavia porcellus* (NC000884.1)

- Después del ensamblaje de 10 genomas, se puede comparar las secuencias reconstruidas a partir de un alineamiento de las secuencias, este fue realizado con BLAST con el programa Gcview webserver. Genoma de referencia NC\_000884.1.
- El ensamblaje de genomas mitocondriales presenta regiones no definidas (presencias de bases “N”)
- Región repetitiva y D-Loop con baja profundidad del gen de ARNt-Leucina e inicio de la región de D-Loop.

No todas las secuencias pueden ensamblarse por lo que a aquellas secuencias únicas que permanecen sin ensamblar se les denomina “singletons”



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

## Validación de ensamblaje

Una vez realizada la reconstrucción del genoma, esta debe enfrentar una serie de validaciones, que están delimitadas por tres métricas:

- Métricas de contigüidad, se analiza el número de *contigs* (secuencias parcialmente ensambladas), en relación al tamaño del genoma.
- Métricas de coherencia, permiten comprobar que los ensamblajes corresponden a las secuencias iniciales utilizadas, para ello se realiza un nuevo alineamiento de las secuencias al genoma ensamblado.
- Métricas de Integridad, se refieren a la composición de genes, presentes en el ensamblaje, si presenta ausencia o duplicación de genes, es un indicativo de buen ensamblaje.

## Validación de ensamblaje de genoma

Resumen de Calidad de secuenciamiento de genoma mitocondrial con el programa SPAdes.

Código	Métricas de contigüidad				Métricas de coherencia					Métricas de Integridad			
	Fracción del genoma (%)	N50	NG50	LG50	Tamaño total de alineamiento (pb)	Alineamiento (%)	Propiamente pareados	<i>Singletons</i> (%)	Cobertura $\geq 1$ (%)	CDS <u>Obs/Esp</u>	ARnt <u>Obs/Esp</u>	ARNr <u>Obs/Esp</u>	
EEA - Chumbibamba	SA42911	96.851	16533	16533	1	16285	99.92	99.22	0.08	99.94	13/13	22/22	2/2
	SA42912	96.851	16335	16335	1	16285	99.91	99.07	0.08	99.76	13/13	22/22	2/2
	SA42913	96.839	16432	16432	1	16363	99.89	98.93	0.09	99.88	13/13	21/22 *	2/2
	SA42914	97.268	15696	15696	1	16361	99.73	99.06	0.27	99.88	13/13	22/22	2/2
Silvestres	SA42976	97.399	16393	16393	1	16393	99.7	98.27	0.29	99.94	13/13	21/22 *	2/2
	SA42977	97.393	15265	15265	1	16502	99.68	98.17	0.28	99.88	13/13	22/22	2/2
	SA42978	98.048	11915	11915	1	16575	99.63	98.92	0.34	99.88	13/13	22/22	2/2
EE- La Molina	SA42979	98.81	16747	16747	1	16606	99.78	97.63	0.22	99.3	13/13	22/22	2/2
La Molina	SA42980	97.399	16414	16414	1	16398	99.5	98.05	0.47	99.85	13/13	22/22	2/2
	SA42981	97.56	16402	16402	1	16402	99.48	98.08	0.52	99.94	13/13	22/22	2/2

\* Ausencia del gen de *ARnt-Leucina*

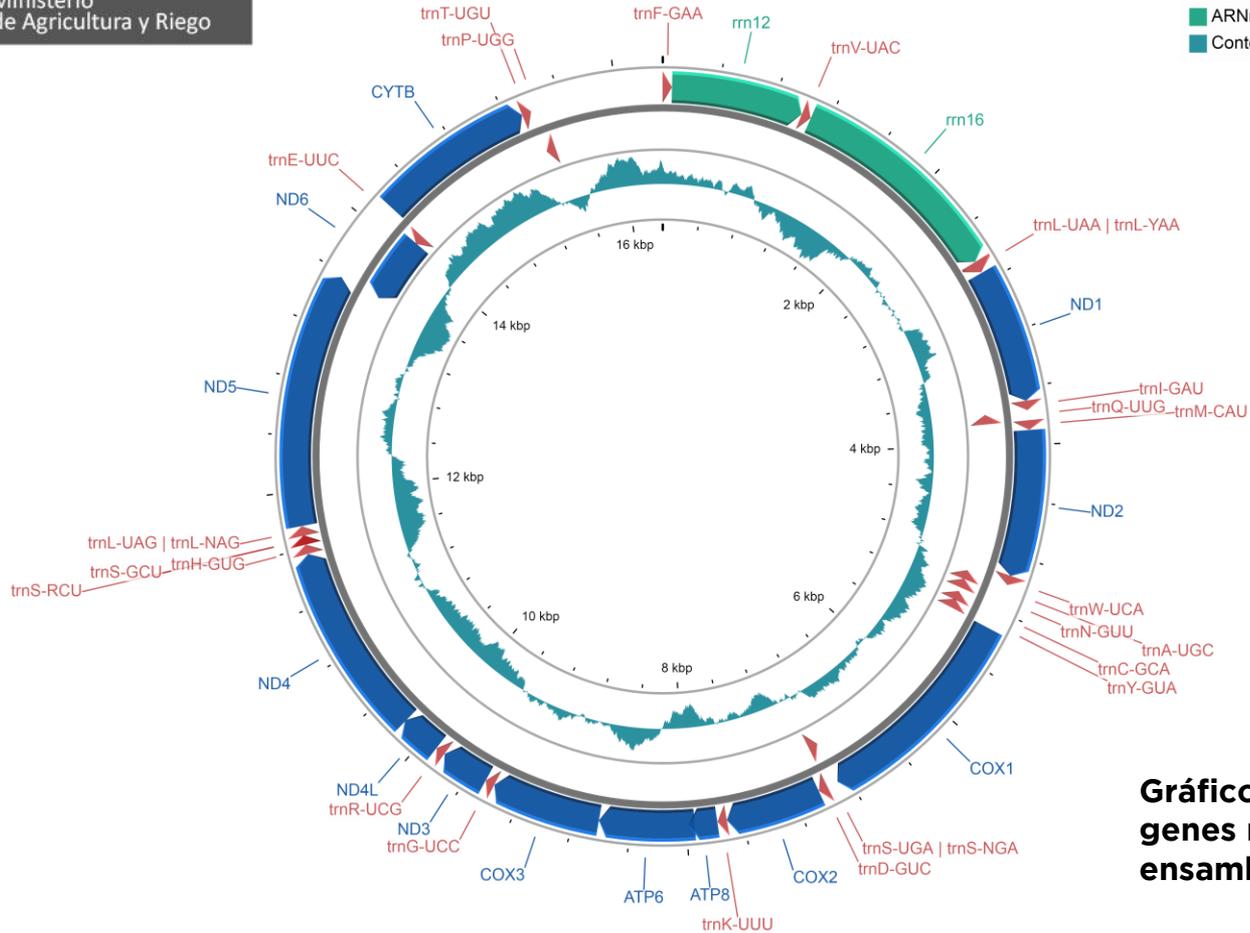
Se realizó validación con **QUAST** y con el servidor web de **MITOSwebserver**





PERÚ

Ministerio de Agricultura y Riego



**Gráfico de ubicación de genes mitocondriales ensamblados**



PERÚ

Ministerio  
de Agricultura y Riego



*Instituto Nacional de Innovación Agraria*

# Análisis filogenéticos

- La obtención de secuencias genómicas (mitocondriales) ensamblados por SPADes, permiten la inferencia del grado de relación filogenética de las especies.
- Se construyó un árbol filogenético con el programa RaxML, basado en la reconstrucción de Máxima Verosimilitud bajo el modelo de sustitución nucleotídica de **GTRgamma**, con bootstrap de 1000 iteraciones.

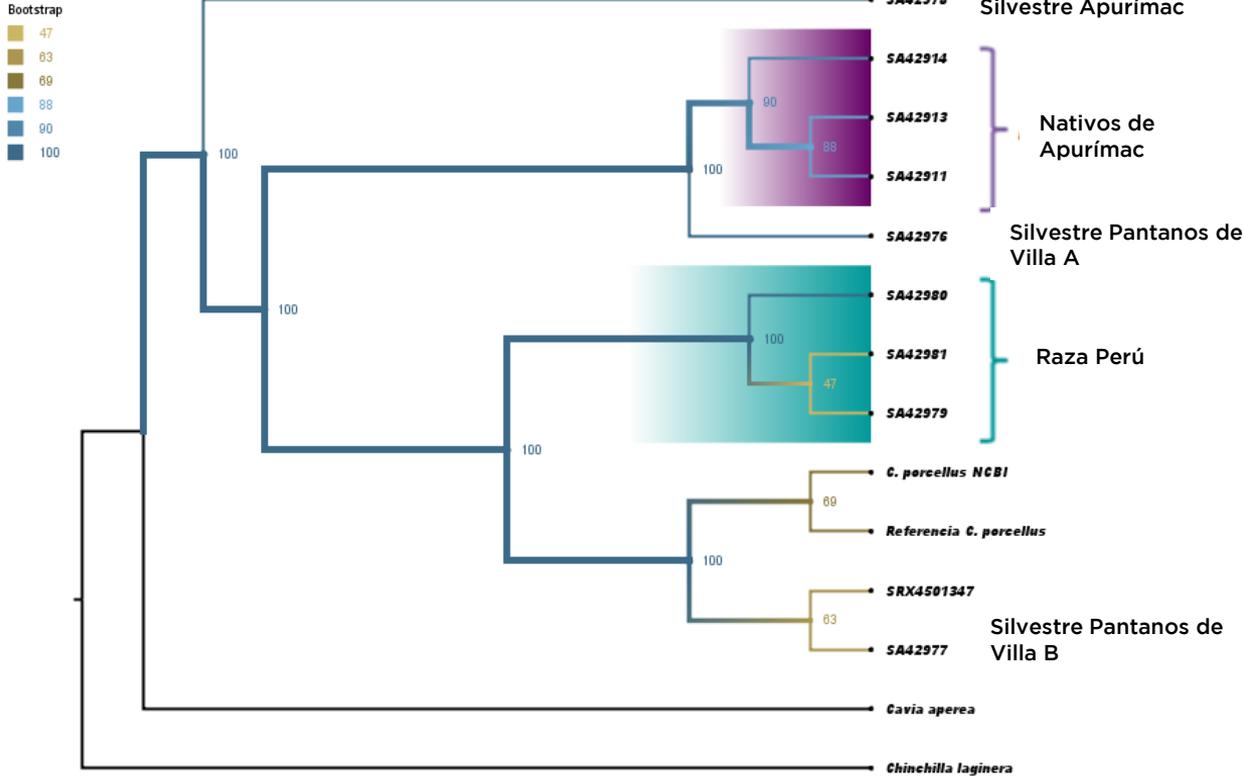


PERÚ

Ministerio de Agricultura y Riego



Instituto Nacional de Innovación Agraria



- El cuy doméstico de Apurímac, se encuentra más relacionado al cuy silvestre de Pantanos de Villa A que al cuy de Raza Perú.
- El cuy de raza Perú se encuentra más relacionado a los cuyes de referencia del NCBI y al cuy silvestre de Pantanos de Villa B.
- El cuy silvestre de Apurímac esta alejado de los silvestres de Pantanos de Villa, puede ser una subespecie de *Cavia tschudii*.

Dendograma de Máxima Verosimilitud basado en los ensamblajes de genomas mitocondriales y genomas representativos y de referencia



PERÚ

Ministerio  
de Agricultura y Riego



Instituto Nacional de Innovación Agraria

**Investigadores Responsables:**  
Blga. Fredesvinda Carrillo Castillo.  
Ing. Eudocio Veli Rivera.

**Equipo Técnico**  
Blga. Claudia Esther Yalta Macedo.  
Blga. Wendy Acuña Rodríguez.  
Blgo. Rolando James Valladares  
Delgado.  
Bach. Lenin Chumbe Nolasco.  
Bach. Juan Torres Chuquillanqui.

**Tesistas:**  
Bach. Roy Sarmiento Sulca.  
Bach. Francisco Ascue Orosco.





PERÚ

Ministerio  
de Agricultura y Riego



*Instituto Nacional de Innovación Agraria*

EL PERÚ PRIMERO

Contacto: [fcarrillo@inia.gob.pe](mailto:fcarrillo@inia.gob.pe)