



Abschließender Sachstandsbericht
Leibniz-Wettbewerb

BRIDGE: Biodiversity informatics to bridge the gap from genome
information to educated utilization of genetic diversity hosted in
Genebanks

Antragsnummer: K180/2014

Aktenzeichen: SAW-2015-IPK-3-BRIDGE

Berichtszeitraum: 01.05.2015- 31.12.2018

Federführendes Leibniz-Institut: Leibniz-Institut für Pflanzengenetik und
Kulturpflanzenforschung (IPK), Gatersleben

Projektleiter/in:
Prof. Dr. Nils Stein

Inhalt

1.	Zielerreichung und Umsetzung der Meilensteine	2
2.	Aktivitäten und Hindernisse.....	2
3.	Ergebnisse und Erfolge.....	3
4.	Chancengleichheit.....	5
5.	Qualitätssicherung.....	5
6.	Zusätzliche eigene Ressourcen.....	6
7.	Strukturen und Kooperation	6
8.	Ausblick.....	6

1. Zielerreichung und Umsetzung der Meilensteine

Erläutern Sie kurz den Stand der Umsetzung der wichtigsten Ziele und Meilensteine Ihres Vorhabens im Berichtszeitraum gemäß Antrag. Erläutern Sie gegebenenfalls die nur teilweise oder nicht erfolgreiche Umsetzung für einzelne Ziele / Meilensteine.

Das übergeordnete Ziel des BRIDGE Projektes war die bessere Integration genotypischer und phänotypischer Daten unter Nutzung innovativer Bioinformatiklösungen sowie die intuitivere Verfügbarmachung der Information als Grundlage für die gezieltere Auswahl genetischer Ressourcen für Forschung und Züchtung.

Alle zentralen Meilensteine wurden erreicht oder übertroffen:

- 1) Genotypendaten (GBS) wurden für alle IPK Gersteakzessionen sowie für insgesamt 3.500 Gersteakzessionen aus zusammen vier weiteren internationalen Genbanken erhoben und analysiert.
- 2) Alle historischen Charakterisierungs- und Evaluierungsdaten für ca. 30 phänotypische Merkmale wurden vervollständigend digitalisiert und für die gemeinsame Analyse mit den Genotypinformationen aufbereitet, verarbeitet und in GBIS und EURISCO importiert.
- 3) Bestäubungseignung wurde in einem Jahr im Feld an einer Auswahl von Akzessionen bonitiert. Desweiteren wurden von >8.000 Akzessionen hochheritable Ähren- und Saatgutmerkmale erfasst und in das BRIDGE Warehouse eingespielt.
- 4) Anstelle von Exome-capture Resequenzierung von 100 Genotypen wurden 300 Genotypen (200 H.v. ssp. Vulgare, 100 H.v.ssp. spontaneum) zu 3-fach Redundanz vollständig sequenziert. Für zwei Genotypen wurde als Beitrag zum Gerste Pan-Genomprojekt ein Chromosomen-ähnliches Genomassembly erstellt.
- 5) Alle Daten wurden in dem, durch das IPK LIMS System unterstützte BRIDGE Warehouse zusammengeführt und mittels neuer Abfragemasken der Öffentlichkeit zur Verfügung gestellt.
- 6) Mehrere Vorträge zum BRIDGE Projekt, Genebank Genomics und Genebank Bioinformatics wurden an der MLU Halle gehalten.

2. Aktivitäten und Hindernisse

Schildern Sie die Arbeiten und Aktivitäten der Projektpartner im Berichtszeitraum. Gehen Sie auch auf möglicherweise aufgetretene wissenschaftliche Fehlschläge und Probleme in der Vorhabenorganisation oder technischen Durchführung ein, die zu Verzögerungen, zu Abweichungen vom ursprünglichen Konzept oder zum Nichterreichen der Zielsetzungen geführt haben.

Sachstandsbericht Leibniz-Wettbewerb [BRIDGE: Biodiversity informatics to bridge the gap from genome information to educated utilization of genetic diversity hosted in Genebanks] [2015-2018]

- 1) Im Rahmen des Projektes galt es die gesamte Gerstenkollektion (>20.000 Akzessionen: 3.902 Winter- und 15.372 Sommergersten, 1.195 Wildgersten, 1.111 Muster einer Core Collection) aus dem Samenkühlager zu entnehmen. Dies erfolgte von April 2015 bis Mai 2017.
- 2) Von jeder Akzession wurde eine Pflanze kultiviert, DNA gewonnen und SSD Saatgut geerntet. Letzteres war im Antrag nicht vorgesehen und wurde durch zusätzlichen Aufwand erreicht. Nach Ernte und Aufarbeitung wurde das Saatgut vakuumiert und bei -18°C gelagert.
- 3) Von allen >20,000 Akzessionen wurden zwei Blattproben (A und B Probe) genommen und DNA isoliert.
- 4) Die ursprünglich in drei Jahren geplante Bonitur der Antherenextrusion als Merkmal für gute Fremdbestäuberleistung wurde nur in einem Jahr erhoben, weil bereits in einem Jahr gute Heritabilitäten erreicht wurden. Als Kompensation wurde zum Zeitpunkt der Saatguternte der Einzelpflanzen von ca. 9000 Akzessionen phänotypische Daten von Infloreszenzen und Samen erhoben.
- 5) Alle A DNA-Proben wurden mittels GBS genotypisiert.
- 6) Von 100 Akzessionen wurden je 10 Nachkommen genotypisiert, um die Reinheit von Sammlungsmustern zu dokumentieren.
- 7) Die Gesamtdiversitätsmatrix wurde in die Projektdatenbank eingespielt und die Sequenz-Rohdaten zum öffentlichen EMBL ENA Sequenzarchiv übermittelt.
- 8) Historische Boniturdaten (seit 1946) wurden für das Genbankinformationssystem konsolidiert, mit Hilfe statistischer und manueller Analysen kuratiert und importiert.
- 9) Für drei Merkmale (Blühzeitpunkt, TKG, Pflanzenhöhe) wurden die historischen Daten mit weiteren statistischen Werkzeugen aufbereitet.
- 10) Zu allen Akzessionen wurden Passportdaten zur Verfügung gestellt und integriert.
- 11) Sämtliche Projektdaten (Aussaat, DNA Extraktion, Phänotypisierung, Genotypisierung) wurden im IPK LIMS erfasst und dem BRIDGE Datawarehouse zugeführt.
- 12) Die Integration von Passport- und Sequenzdaten ermöglichte es erstmals, Aussagen über Herkunft und Spezies der Akzessionen zu treffen, die unabhängig von der Passportdatenlage sind. Dies eröffnet neue Ansätze in der Datenkuration von Passportdaten in GBIS.
- 13) Es wurde eine generische Schnittstelle zwischen Bridge Data Warehouse und dem Bestellsystem des Genbankinformationssystems GBIS entwickelt.
- 14) Es wurde eine Weboberfläche (Webportal) zur Visualisierung und Analyse der Projektdaten entwickelt.

Alle Arbeiten wurden ohne substantielle Verzögerungen gemäß Planung durchgeführt. Aufgrund technischen Fortschritts in der DNA Sequenzierung konnten Ziele übertroffen werden. Der einzige Bereich, in dem Verzögerungen entstanden und nur die Grundziele erreicht werden konnten ist der Bereich Bioinformatik mit Erstellung des Webportals, Datenbank-Programmierung sowie der Programmierung von Abfrage-Use-cases. Die hier entstandenen Verzögerungen sind Schwierigkeiten bei der Personalgewinnung sowie einer teilweise zu optimistischen Meilensteinplanung geschuldet.

3. Ergebnisse und Erfolge

Stellen Sie zentrale Ergebnisse und Erfolge in den Bereichen Forschung (Publikationen, abgeschlossene Qualifizierungsarbeiten, Drittmittelinwerbung, wissenschaftliche Veranstaltungen, etc.) und Transfer (Beratung, Technologietransfer, Presse/Öffentlichkeitsarbeit) dar.

Ergebnisse:

Homogenisierung des Materialmanagements und der Zugriffsschnittstellen

- a) Etablierung einer konsistenten Datenspeicherung durch einheitliche Materialablage im IPK-LIMS und der daraus resultierenden eindeutigen Material-ID;
- b) Alle Daten sind einfach durch Exportfunktionen, SQL-Abfragen, Selektionen abfragbar und für andere Anwendungen z.B. das Erstellen von Webseiten nutzbar;

Sachstandsbericht Leibniz-Wettbewerb [BRIDGE: Biodiversity informatics to bridge the gap from genome information to educated utilization of genetic diversity hosted in Genebanks] [2015-2018]

- c) periodischer Datenabgleich zwischen IPK-LIMS und GBIS zur Unterstützung der kontinuierlichen Kuratierung des GBIS-Datenbestandes.

Genotypisierung/Genebank genomics

- d) genome-weite Markerdaten (> 300,000 SNPs) für fast alle Gersteakzessionen der bundeszentralen Ex-Situ-Genbank am IPK Gatersleben populationsgenetische Analysen zu Strukturierung der Sammlung nach geographischer Herkunft, Zeiligkeit und annueller Wuchsform;
- e) Identifizierung von Duplikatgruppen: Etwa 30 % der Akzessionen sind potenzielle Duplikate;
- f) Entwicklung einer Kernsammlung von 1000 Akzessionen für Assoziationskartierung;
- g) Assoziationskartierung mithilfe historischer Evaluierungsdaten;
- h) Assoziationskartierung von Virus-Resistenz in Zusammenarbeit mit dem JKI auf Basis von Resistenzdaten für >1,800 Wintergerstenakzessionen;
- i) *Allele mining* für Kandidatengenen in einer Kernsammlung von 200 Genotypen;
- j) hochqualitative Referenzgenomsequenzen für zwei Genotypen (Wintergerstensorte Igri und Landrasse HOR 7552) als Beitrag zum vom IPK koordinierten Gerste-Pangenom-Projekt.

Kuratierung und Digitalisierung historischer Charakterisierungs-/Evaluierungsdaten

- k) Aufarbeitung und Konsolidierung von ~1 Million phänotypischen Datenpunkten zu Genbankakzessionen, sowie Import in GBIS und EURISCO,
- l) unabhängige Validierung von Passportdaten auf Basis der gewonnenen Genotypinformationen;
- m) Schnittstelle zwischen dem bestehenden Genbankinformationssystem (GBIS) und dem neuen BRIDGE Datawarehouse.

BRIDGE Datawarehouse / Webportal (<https://bridge.ipk-gatersleben.de>) / Funktionen

- n) Alle im Rahmen des BRIDGE Projektes genutzten Ursprungsdaten sowie die im Projektverlauf generierten wissenschaftlichen Ergebnisdaten wurden in aufbereiteter Form im zugehörigen Projekt-Datawarehouse hinterlegt und sind über das neu entwickelte Webportal zugänglich;
- o) Webportal für kombinierte Suche nach Akzessionen anhand von Passportdaten und den im Projektverlauf gewonnenen phänotypischen Daten. Das Suchergebnis ist als individuelle benannte Kollektion direkt im Webportal speicherbar sowie als CSV-Datei exportierbar;
- p) Markermatrizen visualisierbar über den integrierten SNP-Browser und direkt im Standarddateiformat für Gensequenzvariationen "Variant Call Format" (VCF) exportierbar;
- q) Visuelle Darstellung der geographischen Fundort-Koordinaten;
- r) Möglichkeit zur Verwaltung persönlich ausgewählter Kollektionen von Akzessionen;
- s) Transferfunktion an das Genbankinformationssystem des IPK (GBIS) für ausgewählte Akzessionen zur unkomplizierten Bestellung über GBIS/I bestellt werden;

Erfolge:

a) Publikationen:

- 1) González M.Y. et al. (2018) Scientific Data 5:180278 DOI: 10.1038/sdata.2018.278.
- 2) González, M.Y. et al. (2018) TAG doi: 10.1007/s00122-018-3129-z.
- 3) Milner, S. G. et al. (2019) Nature Genetics 51:319-326.
- 4) Mascher, M. et al. (2019) Nature Genetics 51:1076-1081.

b) abgeschlossene Qualifizierungsarbeiten:

keine

c) Drittmittelinwerbung:

- ICARDA Projekt, bilaterales Projekt IPK-ICARDA, Förderung durch Crop Trust
- BEANadapt, Kooperationsprojekt im Rahmen des ERACAPS Programms, Förderung des deutschen Teilprojektes durch DFG
- G2P-SOL, internationales Kooperationsprojekt, Förderung EU H2020
- Genbank2.0, nationales Kooperationsprojekt, Förderung BMBF
- SHAPE, nationales Kooperationsprojekt, Förderung BMBF
- AGENT, EU-Kooperationsprojekt, Koordination IPK, Antrag in der Begutachtung

d) wissenschaftliche Veranstaltungen (Poster/Vorträge):

Sachstandsbericht Leibniz-Wettbewerb [BRIDGE: Biodiversity informatics to bridge the gap from genome information to educated utilization of genetic diversity hosted in Genebanks] [2015-2018]

37 Vorträge und Poster auf nationalen und internationalen Tagungen

e) **Beratung, Technologietransfer:**

keine

f) **Presse/Öffentlichkeitsarbeit:**

- Diverse Pressemitteilungen IPK ([PM 22.04.2015.](#), [PM 21/2018](#), [PR 21/2018](#), [PM 19/2019](#), [PR 19/2019](#))
- Plant2030 Video ([SHAPE](#))

4. Chancengleichheit

Gehen Sie hier kurz auf Ihre Initiativen und Maßnahmen zur Sicherung der Chancengleichheit, insbesondere bei der Personalentwicklung und Personalgewinnung, ein.

Das IPK fördert aktiv die Gleichstellung von Frauen und Männern und versteht die Förderung von Gleichstellung als eine Leitungsaufgabe. In der Personalgewinnung zielt die passgenaue Ansprache qualifizierter Wissenschaftlerinnen auf die Erhöhung des Frauenanteils – wie das eigens entwickelte Wettbewerbsverfahren zur Gewinnung von Nachwuchsgruppenleiterinnen. Dieses Ziel wird durch verschiedene strukturelle Maßnahmen flankiert, wie die Entwicklung eines umfassenden Gendermonitorings, beispielsweise die Fortschreibung des noch laufenden Gleichstellungskonzeptes des IPK für die nächste Periode von 2020 bis 2024. Darüber hinaus strebt das IPK die Re-Zertifizierung des Audits „Beruf und Familie“ im Rahmen eines Dialogverfahrens an.

Tabelle 1: Anzahl der beschäftigten Personen in den Beschäftigtengruppen nach Geschlecht

Beschäftigtengruppe	männlich	weiblich
Wissenschaftsstützendes Personal	0	4
Studentische und Wissenschaftliche Hilfskräfte	2	2
Wissenschaftliches Personal	3	3
gesamt	5	9

5. Qualitätssicherung

Beschreiben Sie kurz Ihre Maßnahmen zur Qualitätssicherung, insbesondere hinsichtlich der Einhaltung guter wissenschaftlicher Praxis sowie bei der Verfügbarmachung (Open Access) ihrer Forschungsergebnisse.

Am IPK, einschließlich des BRIDGE Projekts, gilt die Einhaltung „guter wissenschaftlicher Praxis“ basierend auf entsprechenden Richtlinien der Deutschen Forschungsgemeinschaft (DFG) und der Leibnizgemeinschaft. Das IPK befindet sich in der Implementierungsphase eines institutsweiten Research Data Management Konzeptes mit dem Ziel der Erreichung der FAIR Standards für alle erhobenen Forschungsdaten.

IPK hat eDAL-PGP (<https://edal-pgp.ipk-gatersleben.de>) als anerkanntes Repository für die Publikation von Datensätzen und deren nachhaltige Nachnutzung etabliert. Mit Hilfe dieses Werkzeugs wurden folgende Datensätze gemäß der Open Data -Standards publiziert: Variant Matrix of wild barley: <http://doi.org/10.5447/IPK/2019/3>; Variant Matrix of global barley diversity panel: <http://doi.org/10.5447/IPK/2018/9>; Characterization & Evaluation data for three agronomic traits: <http://doi.org/10.5447/IPK/2018/10>

Darüber hinaus wurden alle zentralen Roh-Datensätze zur Genotypisierung der IPK Gerstensammlung an das öffentliche DNA-Sequenzdatenarchiv EMBL-ENA eingereicht (Akzessionsnummern: [PRJEB23967](#), [PRJEB24563](#), [PRJEB24627](#), [PRJEB26634](#), [PRJEB26652](#), [PRJEB27184](#)). Hierbei wurden alle Identifikatoren der EMBL-ENA Datenbank den im IPK-LIMS gespeicherten Materialien zugeordnet und Materialien-IDs aus dem IPK-LIMS mit in der EMBL-ENA Datenbank gespeichert, um per bi-direktionaler Verknüpfung die FAIR-Prinzipien zu erfüllen.

6. Zusätzliche eigene Ressourcen

Bitte schätzen Sie die Höhe der Mittel ab, die im Rahmen des Projekts als „In-kind“ Leistungen an ihrem Institut, den beteiligten Leibniz-Instituten und/oder den universitären Kooperationspartnern erbracht wurden. Differenzieren Sie dabei nach Personalmitteln (z.B. in Personenmonaten, wissenschaftliche und nichtwissenschaftliche Mitarbeiter) und Sachmitteln.

Für die Förderformate Leibniz-Professorinnenprogramm und Leibniz-Junior Research Groups geben Sie bitte auch die Höhe der Kofinanzierung durch die Leibniz-Einrichtung an.

Das IPK unterstützte das BRIDGE Projekt über den gesamten Zeitraum mit eigenen Ressourcen in Höhe von 587.933,00 EURO (Tabelle 2).

Tabelle 2: Übersicht „In-kind“ Leistungen des IPK

	Personenmonate (PM)	EURO
Technisches Personal	48	236.500
Wissenschaftliches Personal	46	251.433
Infrastruktur		100.000
Summe	94	587.933

7. Strukturen und Kooperation

Beschreiben Sie die konkrete Ausgestaltung bestehender Kooperationen im Berichtszeitraum inkl. Möglicher Hindernisse und Herausforderungen. Ergänzen Sie ggf. Veränderungen der Governance Ihres Vorhabens und/oder die Aufnahme neuer Partner bzw. Kooperationsbeziehungen. Sind im Berichtszeitraum neue Kooperationsbeziehungen zu institutionellen Partnern aufgenommen worden, erläutern Sie bitte kurz den wissenschaftlichen Nutzen für Ihr Vorhaben.

Sollten die Veränderungen Zusatzvereinbarungen zur oder Änderungen der bestehenden Kooperationsvereinbarung mit sich gebracht haben, fügen Sie diesem Bericht bitte entsprechende Kopien bei.

Die ursprüngliche Planung zur Durchführung des Projektes basierte in erster Linie auf intensiver IPK-interner Kooperation der interdisziplinären Arbeitsgruppen des Instituts, insbesondere der Abteilungen Genbank und Züchtungsforschung. Darüberhinaus war eine Kooperation mit der Martin-Luther-Universität Halle geplant, insbesondere in der Masterausbildung im Bereich Pflanzenzüchtung und Bioinformatik. Die IPK-interne Kooperation funktionierte hervorragend. Es wurden neue Maßstäbe für die institutsweite, interdisziplinäre Kommunikation und Kooperation gesetzt. Die Kooperation mit der Universität Halle konnte nicht die gewünschte Intensität erreichen.

Das Projekt BRIDGE hat dazu beigetragen hervorragende internationale Kooperationen zu vier weiteren internationalen Genbanken (Vavilov Institut Russland, ICARDA Marokko, CAAS Genbank China, Genbank Schweiz) zu etablieren was mit die Grundlage für einen IPK-koordinierten und eingereichten EU H2020 Antrag bildete.

Während der Projektlaufzeit wurden drei Projekte auf Basis des BRIDGE Projekt Konzeptes mit IPK-Beteiligung gestartet (Tabelle 3).

Tabelle 3: Drittmittelprojekte eingeworben während der Laufzeit des BRIDGE Projektes

Projektname	IPK AG	Förderprogramm	Förderagentur
BEANadapt	GED	ERACAPS	DFG
G2P-SOL	GGR, RGR, TEN, DOK	H2020	EU
Genbank2.0	QG, DG, GGR, BIT, DOK	Plantbiot. f. Breed. Res.	BMBF

8. Ausblick

Beschreiben Sie die wichtigsten zukünftigen Forschungsfragen und -felder.

Das Projekt BRIDGE hat wesentliche Weichen für die Weiterentwicklung strategischer Forschungsgebiete des IPK gestellt:

Sachstandsbericht Leibniz-Wettbewerb [BRIDGE: Biodiversity informatics to bridge the gap from genome information to educated utilization of genetic diversity hosted in Genebanks] [2015-2018]

- Erschließung des Forschungsfeldes „Genebank genomics“;
- Systematische Etablierung sequenzbasierter Marker als “molekulare Passportdaten” auch für Sammlungen anderer Fruchtarten am IPK (z.B. Weizen, Solanaceen, Bohne);
- Aktives Kuratieren der IPK-Sammlung; Detektion und Behandlung von Duplikaten; Validierung und Korrektur von Passport-Daten;
- Transformation von Genbanken zu sogenannten bio-digitalen Ressourcententren;
- Grundlage eines EU-Antrages zur Charakterisierung von Sammlungen europäischer Genbanken (Antrag AGENT eingereicht in SFS-28-2019 B: Adding value to plant Genetic Resources (GenRes));
- Definition von Kernsammlungen für Phänotypisierung und tiefere Genomanalyse;
- Genotypenauswahl für das Internationale Gersten-Pangenom-Projekt;
- Sequenzierung einer Kernsammlung von 1000 Gerste Genotypen (BMBF-Finanzierung).