

生物種メタボロームモデル・ データベースの構築

有田正規 (理研CSRS)

金谷重彦 (NAIST)

櫻井望 (かずさDNA)

前期までの主要結果

- Bio-MassBankの公開(奈良)

生物種	組織など	研究グループ	データ数
シロイヌナズナ	葉	かずさ DNA 研	870
ミヤコグサ	花弁	かずさ DNA 研	908
キャベツ, 3栽培品種		かずさ DNA 研	1,702
トマト, 12栽培品種	果実	かずさ DNA 研	20,303
ハウレンソウ, 4栽培品種	葉	かずさ DNA 研	17,611
ナンヨウアブラギリ, 4栽培品種	果実	かずさ DNA 研	5,481
タマネギ	食用部	理研、植物科学センター	72
ディクソニア・アンタルクティカ、	葉	かずさ DNA 研	434
ヒメツリガネゴケ	葉	かずさ DNA 研	484
ゼニゴケ, 2種	葉	かずさ DNA 研	1,107

前期までの主要結果

微生物

生物種	組織など	研究グループ	データ数
シアノバクテリア PCC6803	細胞	かずさ DNA 研	124
クラミドモナス	細胞	かずさ DNA 研	146
マイタケ	子実体	かずさ DNA 研	1,424

哺乳動物

生物種	組織など	研究グループ	データ数
ハツカネズミ	3臓器、脂質	中部大学	2,250
ヒト (* 公開準備中)	尿	CEA - Centre d'Etude de Saclay, Gif-sur-Yvette, France	13,306

総計 66,272 データ

前期までの主要結果

- Metabolonote等の公開(かずさ)

METABOLO NOTE

SE1:/S01/M01/D01

Sample Set Information

ID	SE1
Title	Arabidopsis thaliana leaf metabolite analysis
Description	Investigation of Arabidopsis thaliana leaf metabolites. 6 replicated data are examined for each sample.
Authors	Takashi Arai, Ryohei Sakai, Hitaru Enomoto, Ritsuro Sekura, Hideyuki Suzuki, Daisuke Shibata, Kazusa DNA Research Institute
Reference	Direct Submission
Comment	version 5

PGDB The web resource and information related to the species used in this study are available at Plant Genome DataBase Japan (PGDBJ)

Sample Information

ID	S01
Title	Arabidopsis wt leaf
Organization	Arabidopsis thaliana
Organization Name	Arabidopsis thaliana
Organization ID	KOJI2 taxonomy:PT02
Compound ID	
Compound Name	
Preparation	Arabidopsis thaliana Col-0 seeds are sown on pots filled by vermiculite and are grown in an incubator with 16h light/8h dark and 22 degree C condition. After 2 months later, all leaves are harvested.
Comment	[KomicMarket ID] <SBA_13

KOMIC Market The annotation data and quantitative data of the metabolite peaks in this sample are available at KomicMarket

Analytical Method Information

ID	M01
Method and ID	M01
Sample Amount	0.7 mg
Comment	[MassBase ID] M01_CL_05711

MassBase The raw (binary) and near raw (text) files of this analysis are available at MassBase

Analytical Method Details Information

ID	M01
Title	LC-FT-SEC-MS ESI positive method 1
Instrument	Agilent1100 HPLC (Agilent), LTQ-FT (Thermo Fisher Scientific)
Instrument type	LC-FT/SEC-MS
Ionization	ESI
Ion Mode	Positive
Description	Harvested sample is frozen by liquid N2 and resulting powder (100mg) are solved in 300uL 80% methanol is injected into HPLC after 0.2um membrane filter treatment. HPLC conditions: Agilent 1100 (Column: Thermo2000 (4.6 x 250 mm, 5 micrometer) T0020), Solvent: A: 0.2% formic acid in water, B: 100% acetonitrile (A:B 10:90 to 25:75 (10 min), 25:75 to 20:80 (25 min), 20:80 to 10:90 (45 min), 10:90 to 50:50 (15 min), 50:50 to 50:50 (15 min), 50:50 to 25:75 (15 min), 25:75 to 10:90 (15 min), 10:90 to 10:90 (15 min)), Gradient: (8:9:2 to 20% (0.0 to 25.0 min), 20 to 90% (25.0 to 40.0 min), 90% (40.0 to 50.0 min), 2% (50.1 to 57.0 min), Column temp: 30 degree C, Flow rate: 0.3mL/min, P0 (60s), P1 (200ms conditions) Filter 1: FTMS is a column topology for m/z 20000 to 2000 (0.1500 to 2.1500 min), normal scan rate Dep MS/MS Most intense ion from (1), rejected (mass=256.2000;266.0000;284.2000;300.2000;326.2000;344.2000;372.2000;388.2000;416.5000)
Comment of details	

Data Analysis Information

ID	D01
Data Analysis Set ID	D01
Recommended default places of m/z	ITMS 2
Comment	

Bio-MassBank The mass spectrum data are available at Bio-MassBank

NEW KomicMarket TEMPORARY The peak data files written in TugMD format are downloaded by the New KomicMarket temporary web site

Data Analysis Details Information

ID	D01
----	-----

Bio-MassBank

Mass Record: KZSAT0000000024
Arabidopsis wt leaf / SE1_S01_M01_D01_P0016; LC-ESI-ITFT; MS2; [M+H]⁺; RT:9.576756 min

Mass Spectrum

Accession: KZSAT0000000024
PSI000 TITLE: Arabidopsis wt leaf / SE1_S01_M01_D01_P0016; LC-ESI-ITFT; MS2; [M+H]⁺; RT:9.576756 min
DATE: 2019.05.15 (Created: 2011.12.09)
AUTHORS: Takashi Arai, Ryohei Sakai, Hitaru Enomoto, Ritsuro Sekura, Hideyuki Suzuki, Daisuke Shibata, Kazusa DNA Research Institute
LICENSE: CC-BY-SA
PUBLICATION: Direct Submission
COMMENT: [Metabolonote ID] SE1_S01_M01_D01_P0016
COMMENT: [MassBase ID] M01_CL_05711
COMMENT: [KEGG ID] KEGG0000000024
COMMENT: MS_TYPE details is ITMS.

OSNAME: Glucosaminolactone
COMPOUND_CLASS: Natural Product
CHEMFORMULA: C15H21NO10S
CHEXACT_MASS: 437.0841
CHEMFILES: N/A
CHEMPACT: N/A
CHEMLINK: KEGG 0000419
CHEMLINK: KNApSack 000007545 000007545

SPECIES_NAME: Arabidopsis thaliana
SPECIES_WGT: TAXONOMY 3109
SPECIES_SAMPLE: Arabidopsis wt leaf

ACQUISITION: Agilent1100 HPLC (Agilent), LTQ-FT (Thermo Fisher Scientific)
ANALYTICALMETH: FTMS, LC-ESI-ITFT

KEGG Kyoto Encyclopedia of Genes and Genomes

KNpSack Core System
Since 2004.04

前期までの主要結果

- KNApSack Family DB の拡充(奈良)
 - 9,584 代謝物-活性関係
 - 生物活性のオントロジー
 - 食品中の代謝物

"KNApSack" Family

Since 2008.07

KNApSack Metabolomics

3D
Since 2012.11

Core System
Since 2008.04

Search Engine
Since 2008.12

Pocket Search for Functional Species

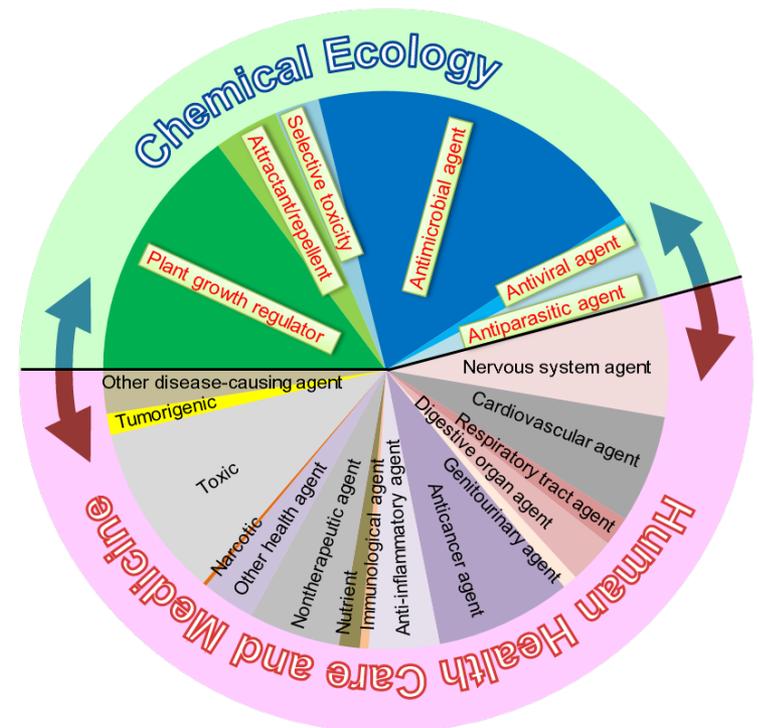
<p>Food & Health</p> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>Lunch Box 食用データベース <small>Since 2008.07</small></p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>DietNavi 病気の予防データベース <small>Since 2012.11</small></p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>FoodProcessor 加工食品データベース <small>Since 2012.11</small></p> </div> <div style="border: 1px solid red; padding: 2px;"> <p>DietDish お食合わせデータベース <small>Since 2012.11</small></p> </div>	<p>Crude Drug</p> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>WorldMap 世界の薬用植物データベース <small>Since 2009.04</small></p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>KAMPO 漢方薬データベース <small>Since 2009.04</small></p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>JAMU IndonesiaHerbデータベース <small>Since 2009.11</small></p> </div> <div style="border: 1px solid red; padding: 2px;"> <p>Tea Pot データベース <small>Since 2011.09</small></p> </div>	<p>Biology</p> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>Biological Activity Natural Activity <small>Since 2011.08</small></p> </div> <div style="border: 1px solid red; padding: 2px;"> <p>Biological Activity Metabolite Activity <small>Since 2013.01</small></p> </div>
---	---	--

<p>Picnic Gene Annotation</p> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>Arabidopsis <small>Since 2008.04</small></p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>Bacillus <small>Since 2008.05</small></p> </div> <div style="border: 1px solid red; padding: 2px;"> <p>Human <small>Since 2009.03</small></p> </div>	<p>Strap Correlation Coefficient</p> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>Arabidopsis <small>Since 2009.09</small></p> </div> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>Bacillus <small>Since 2009.09</small></p> </div>
--	--

<p>Motorcycle Metabolic Pathway</p> <div style="border: 1px solid red; padding: 2px; margin-bottom: 5px;"> <p>代謝データベース <small>Since 2011.08</small></p> </div>	<p>Pickaxe Metalloprotein Database</p> <div style="border: 1px solid red; padding: 2px;"> <p>MetalMine <small>Since 2009.08</small></p> </div>
--	--

<p>Bicycle Algae Metabolic Pathway</p> <div style="border: 1px solid red; padding: 2px;"> <p>代謝データベース <small>Since 2013.09</small></p> </div>

Skewered KNApSack 串刺し検索
Since 2010.10

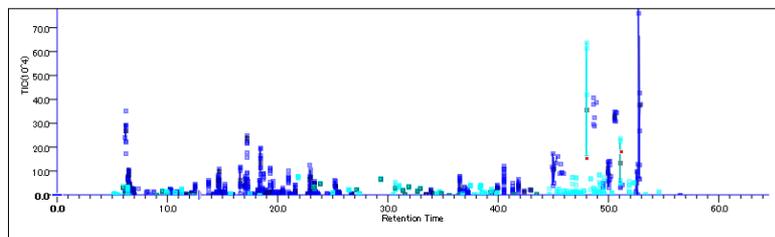


前期までの主要結果

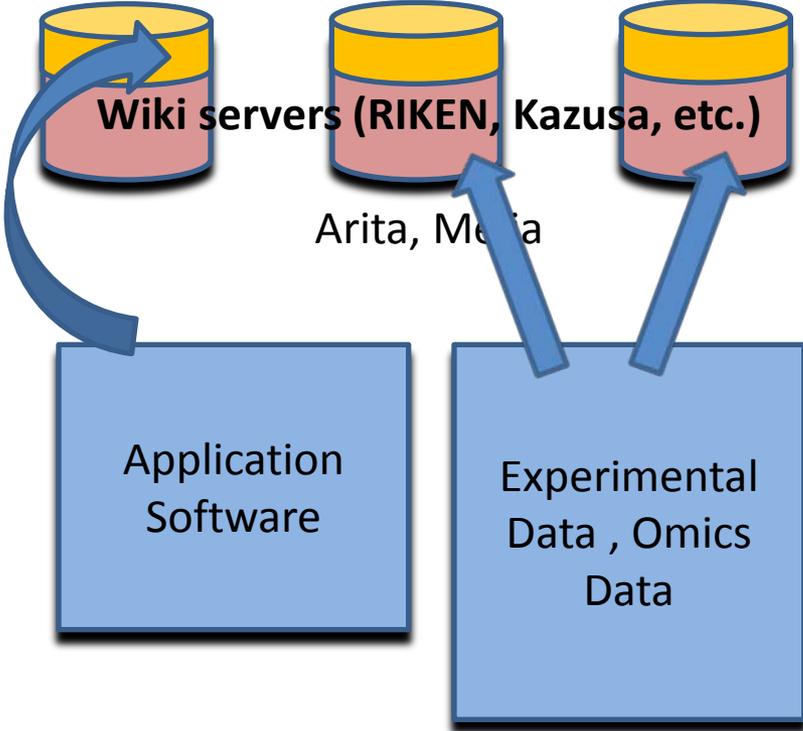
データ共有ソフトの作成 (理研)



Arabidopsis/wt/leaf

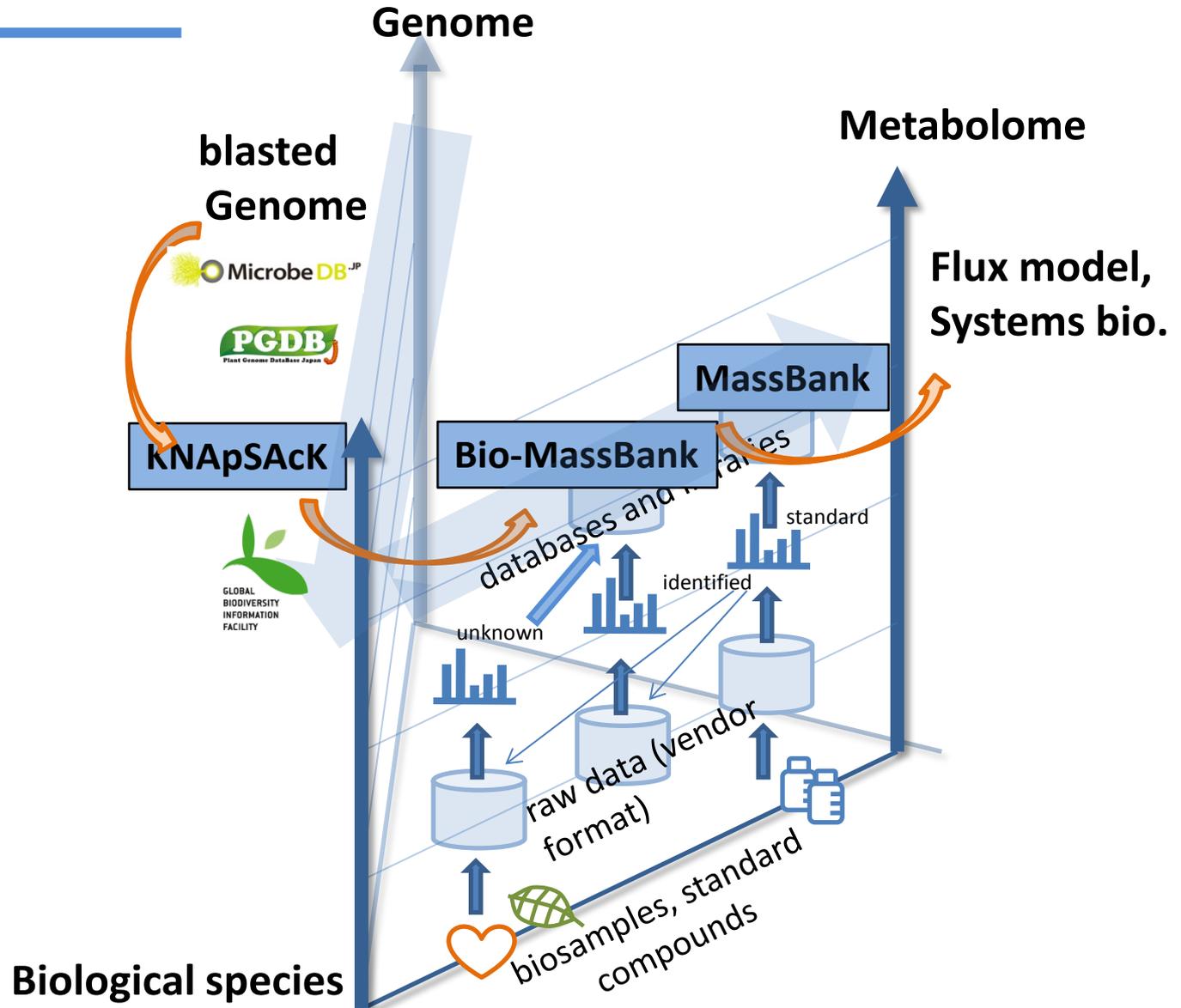


Networking Layer (Akie Mejia)



今回の提案

ゲノム
+ 文献情報
↓
代謝物リスト
↓
代謝物同定
スペクトル
↓
システム生物学



目標

- 化合物情報とスペクトル情報を生物種ごとに整理・公開する

モデル含む20種程度、特に植物

- 活性情報を付した化合物を、代謝マップを通してゲノム情報とリンクする

二次代謝物を中心に、脂質をカバー

- オントロジーと標準フォーマットを整備

アノテーションRDF、整理活性RDFで連携

メタボモデル解説

例： トマト

Asteridae -

Acids, Bases[m]

:

alanine etc.

Solanaceae -

Alkaloids [m]

:

Solanum lycopersicon

tomatine [m]

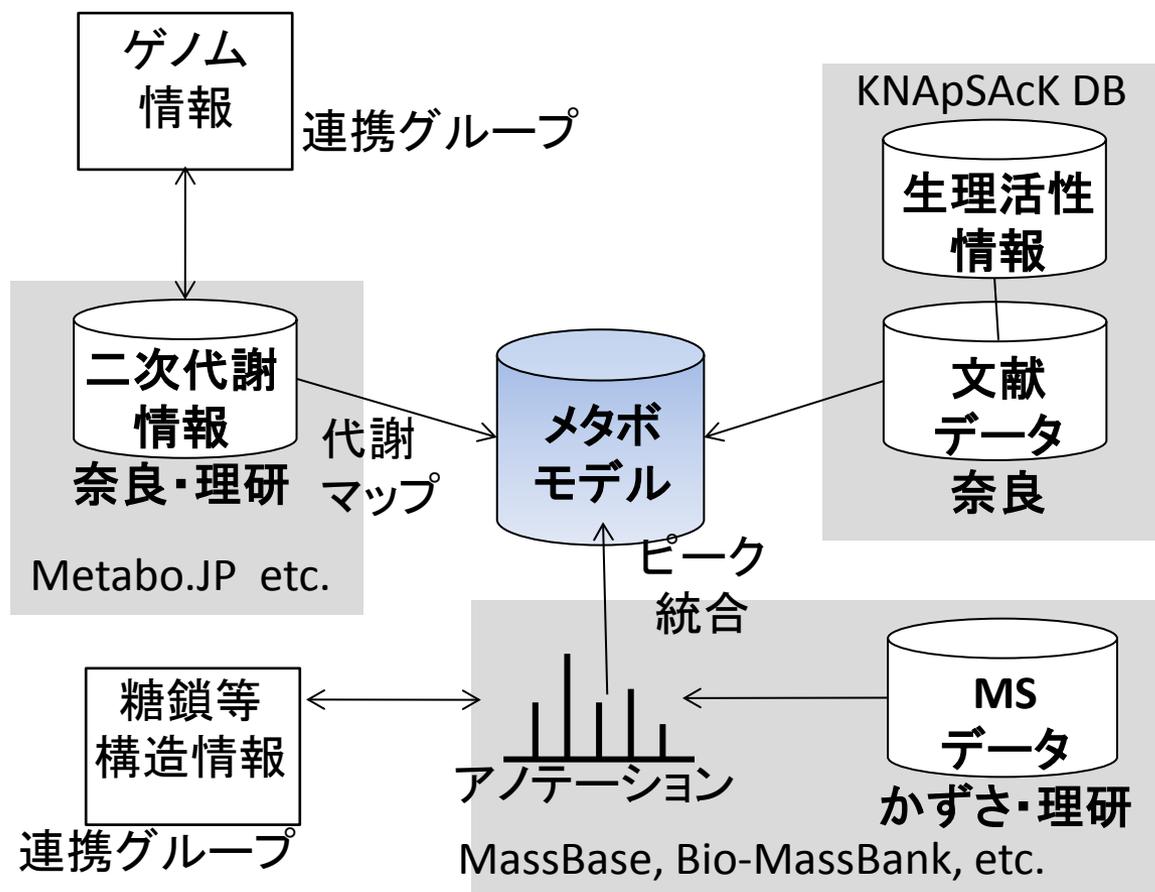
:



検索するDBから、見るDBへ。Bio-MassBank の系統分類版。
各代謝物に、文献に拠る裏付けおよび統合スペクトル情報を付与
物質の生理活性も(分かる場合は)付与 (リンクアウト)

研究体制

旧 MassBank プロジェクト雇用者(鶴岡メンバー)は解散。
KNApSAcKとシステム開発以外は新規雇用。



各グループの研究内容

理研

1. 生物種メタボモデルの構築
2. データ共有基盤ソフトウェアの構築
3. アノテーションソフトウェアの構築

奈良

1. 生物種と代謝物の関係データベースの充実
2. ケミカルエコロジーを中心とした代謝物と生物活性の関係データベース
3. 代謝物の構造情報によるスペクトル情報にもとづいた構造推定法

かずさ

1. メタボロームデータの大規模整備
2. メタデータのRDFによる外部データベースとの連携
3. アノテーションおよびデータ公開加速のための技術開発

目標のまとめ

1. (Bio-)MassBankのスペクトルを拡充・統合し、生物の系統毎に整理、公開
2. KNApSAcKを介して、代謝物とゲノム情報、生理活性をリンク
3. 必要なフォーマット、オントロジー、ツールの作成

