

代謝ネットワークギャップの網羅的遺伝的相互作用解析による補完経路の推定

田中 雄一郎
奈良先端科学技術大学院大学
バイオサイエンス領域 システム微生物学研究室
(森 浩禎 教授)
令和3年01月18日提出

目次

序論.....	3
1.1 頑強性.....	4
1.2 代謝経路が示す頑強性と遺伝的相互作用.....	5
1.3 頑強性を調べる上での課題.....	7
1.4 遺伝的相互作用検出のための二重遺伝子欠失実験.....	10
1.5 接合伝達を利用した二重欠失株作製手法.....	10
1.6 Colony-live について.....	14
1.7 未知の代謝経路がもたらす代謝シミュレーションの偽陰性の 解明に向けたアプローチ.....	17
2 章.....	22
2.1 緒言.....	23
2.1.1 最少培地で偽陰性 (False negative:FN)を示す遺伝子の特定.....	23
2.1.2 最少培地で FN を示す遺伝子を対象とした二重欠失株(Double KnockOut: DKO)解析において発生した問題.....	24
2.2 実験条件.....	26
2.2.1 培地及び化学物質.....	26
2.2.2 菌株及びプラスミド.....	27
2.2.3 一遺伝子欠失株および二重欠失株コロニーの生育観察.....	27
2.2.4 シミュレーション実験に使用したソフトウェアと条件.....	30
2.2.5 最少培地上での DKO 株作成プロトコルの見直し.....	31
2.3 結果.....	33
2.3.1 糖源の違いによる実測値.....	33
2.3.2 シミュレーションと実測との間に食違いのある遺伝子群.....	40
2.3.3 最少培地を用いた二重欠失株作製法の確立.....	42
2.4 考察.....	50
2.4.1 一遺伝子欠失実験及びシミュレーションとの比較.....	50
2.4.2 最少培地を用いた DKO 株作製プロトコルの改変.....	54
3 章.....	57
3.1 緒言.....	58
3.2 手法.....	60
3.2.1 生育のばらつきと落ち込みによるデータの抽出.....	60
3.2.3 欠損値の補填方法.....	61
3.2.4 ϵ スコアの計算.....	61
3.2.5 生育プロファイルに基づいたクラスタリング.....	62
3.2.6 負の遺伝的相互作用検出.....	62
3.3 結果.....	63
3.3.1 生育のばらつきの定量化および評価.....	63
3.3.2 生育の落ち込みの原因と検出および評価.....	65
3.3.4 相同組み換え出来ない箇所の除外と補填方法.....	70
3.3.5 遺伝的相互作用スコアの再定義.....	71

3.3.6	生育度プロファイルに基づいたクラスタリング.....	74
3.3.7	False negative を示す遺伝子と配列類似性および負の遺伝的相互作用を示す遺伝子.....	78
3.4	考察.....	80
文献	82

序論

1.1 細胞の頑強性

バクテリアからヒトに至るまで、地球上で生きる生命は多くの環境変化の中を生きている。生育する場の環境変動のみならず、細胞内の遺伝的要因も環境の変化である。生育環境の温度、湿度、圧力、日光、食糧となり得る物質の存在、その他酵素反応などに必要な金属イオンなどが外的環境であり、紫外線や放射線による遺伝子損傷による遺伝子機能の変化などは内的要因と捉えられる。これら全ての変動に対して、生命体は、その変動による攪乱を生き延びる術を持つ。攪乱に対して生命は何らかの方法で対処し、その危機を乗り越える。これをロバスト性(頑強性)と言う(H Kitano., 2001)。

環境変動に対する生命の生存戦略機構研究の歴史は長い。1962年にイタリアの Ritossa F.は、25°C で生育を続ける *Drosophila busckii* のサナギを30°C30分間で処理を施すと、染色体の特定の場所にパフが観察されることを報告し(Ritossa F., 1962)、その後の Heat shock 応答、Heat shock タンパク質による急激な熱変動による細胞内変化及びその変化を生き延びる機構解明の分野が立ち上がった。また、熱ショックはタンパク質の変性を引き起こしタンパク質本来の機能を失うことで細胞内の機能が一時的に阻害されることがわかり、その後変性タンパク質の感知システムやリサイクルの研究で、京都大学の森和俊博士がノーベル賞候補に、東京工大の大隅良典栄博士が2016年にノーベル医学・生理学賞を受賞されたことは記憶に新しい。一方、熱ショックタンパク質が、個々の細胞が持つ遺伝子変異による影響を緩和するバッファの働きをすることが明らかにされ、複数の遺伝子産物相互作用による表現型調節の存在が示された(Rutherford and Lindquist., 1998)。

このように半世紀以上にわたって、環境変動ストレスを細胞が如何にして自身を守り生存を続けるかの研究が続いてきた。一方、1990年代のゲノムプロジェクトに始まった網羅的実験研究リソースの拡張、急速な技術革新による膨大なデータ蓄積及び理論を基にした実験が容易になってきた背景から、生命をシステムとして捉えようとするシステム生物学が21世紀の生物学として注目を浴びたのが1990年代の後半である。その研究対象の一つとして大きく取り上げられているのが細胞のロバスト性解明である(Bernhard Ø. Palsson., 2015)。

複数の遺伝子による相互作用解析の歴史も同様に古い。例えば合成致死解析やサブ

レッサー変異解析と呼ばれる方法による解析である (P Novick et al., 1989)。ある遺伝子の変異のみでは表現型が出ないが、さらにもう一つの遺伝子に変異が入った場合に致死になる組合せを探索し、遺伝子間の相互作用を解析するものが合成致死解析である。またある遺伝子変異による表現型の変化を、もとに戻す新たな変異を探索するのが、サプレッサー変異解析である。サプレッサー変異に関しては、これまで変異株探索と同様に、ランダムな変異を導入し表現型で選択を行う方法で研究が進められてきた。一方、ゲノム研究進展の成果として、多くのモデル生物に対して全遺伝子をカバーする実験研究リソースの拡充が図られ、網羅的な遺伝子欠失実験によるサプレッサー変異を対象とした研究も系統的に進めることが可能となった。

奈良先端科学技術大学院大学システム微生物学研究室では、大腸菌というモデル生物のゲノムプロジェクトを完成し、ゲノムアノテーション会議を国際協力を進め、そして全遺伝子を対象として実験リソースを整備したグループである (Baba et al., 2006)。これらの実験リソースは、世界中の大腸菌を用いる研究の世界標準として世界中に浸透している。

1.2 代謝経路が示す頑強性と遺伝的相互作用

1997年に日米の大腸菌ゲノムプロジェクトにより、大腸菌K-12株のMG1655とW3110の2種類の全ゲノム構造が明らかになった (Blattner et al., 1997; Yamamoto et al., 1997)。これにより、系統的な解析を可能にする遺伝子クローンや欠失株などの網羅的な実験リソース構築のための準備が整った。その後遺伝子クローン及び欠失株ライブラリーの構築が始められ、一遺伝子欠失株ライブラリーのKeio collection (Baba et al., 2006) 及び全ORF遺伝子プラスミドクローンのASKA library (Kitagawa et al., 2005) が公開された。

Keio collectionの作製の際に、野生株の持つ約4,300のタンパク質をコードする遺伝子のうち約300遺伝子が栄養培地であるLB培地で欠失することができない必須遺伝子であることが示された (Baba et al., 2006)。また、2006年当時の遺伝子アノテーションにおいて翻訳及び細胞分裂に関わる遺伝子における必須遺伝子の数がそれぞれ0.45%と0.56%と、アミノ酸・核酸・補酵素関連遺伝子を合わせた0.08%と比較して

大きいことが分かった。このことは、LB培地では代謝物を合成せずとも培地から供給される栄養源によって生育が可能であることを示していると考えられる。

一方、LB培地上で作製されたKeio collectionを用いてM9最少培地での必須性を全遺伝子について調べることは、M9最少培地で欠失を行う作業を再度行う必要があり現状では困難であるが、少なくともLB培地で取得できた欠失株がM9最少培地で生育可能かどうかは調べるのが可能である。表1から分かるように、LBで必須である翻訳や細胞分裂に分類される遺伝子群については、M9 Glucose培地で致死性を示す遺伝子の追加は見られない。しかし、アミノ酸・核酸・補酵素などの生合成経路に関連する遺伝子群においては、必須遺伝子の数がアミノ酸関連で0.02%から0.15%、核酸関連で0.13%から0.37%、そして補酵素では0.20%から0.38%と、M9培地で大きく増加している。この事実は、酵素遺伝子群の必須性に関しては、培地環境から摂取可能な状態で調べることは不十分であり、代謝経路の頑強性を調べる上でも重要である。

表1. Keio collection構築時に明らかとなった遺伝子の必須性

	必須遺伝子		合計	全遺伝子	全遺伝子における割合(%)	
	LB	M9			LB	LB+M9
translation	77	0	77	170	0.45	0.45
transcription	15	3	18	288	0.05	0.06
replication	15	2	17	237	0.06	0.07
cel division	19	0	19	34	0.56	0.56
post-translational modification	9	2	11	133	0.07	0.08
cell enverope	39	0	39	219	0.18	0.18
cell motility	11	0	11	151	0.07	0.07
inorganic ion transport	1	9	10	253	0	0.04
signal trnsduction	6	0	6	187	0.03	0.03
energy	6	2	8	281	0.02	0.03
carbohydrate transport	8	1	9	355	0.02	0.03
amino acid*	7	52	59	390	0.02	0.15
Nucleotide*	11	20	31	83	0.13	0.37
Coenzyme*	29	25	54	144	0.2	0.38
lipid	24	0	24	96	0.25	0.25
secondary metabolism	5	0	5	95	0.05	0.05
general function	15	4	19	448	0.03	0.04
unknown	18	1	19	447	0.04	0.04
no COG	20	12	32	1214	0.02	0.03

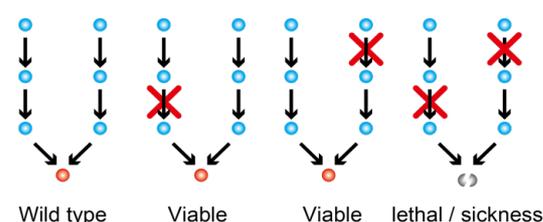
*赤字はM9培地で致死性を示す遺伝子群が多い分類

1.3 頑強性を調べる上での課題

細胞が生存に必須な代謝産物を合成する経路が複数存在すると、一方の経路を遺伝子欠失で遮断しても、他方の経路による合成が可能のため、通常細胞は生存を続けることが可能と考えられる。図1に示すように、同一の代謝物質を合成する経路が2

つであった場合を想定すると、並行経路の存在を同定するには同時にそれら2つの経路を遮断することにより可能になる。これは、冗長な経路のみならず、アイソザイムの様に、同一の反応を行う酵素が複数存在する例が挙げられる。3つ以上並行経路が存在する場合は更なる遺伝子欠失を行う必要が考えられるが、組み合わせの数が膨大になり現状では現実的な時間内で実験を行う事が出来ない為、以降は2つの経路である場合のみで話を進めていく。

A) 二重欠失株による解析



B) 欠失株および酵素阻害剤による解析

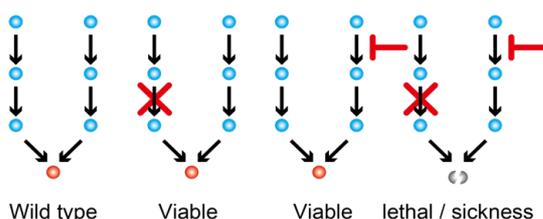


図1 遺伝子欠失株および酵素阻害剤による遺伝的相互作用解析

青および赤の円は代謝産物を示し、矢印は酵素反応を意味する。赤は生育に必須な代謝産物であることを示す。×印は遺伝子欠失を、Tの印は阻害剤による酵素反応阻害を示す。A) 二重欠失による2酵素の同時破壊。B) 一遺伝子欠失に対して、他のステップの化学物質による反応阻害。ともに2つの酵素反応を同時に阻害することで、それら二つの酵素の機能的相補関係の解析が可能である。

図1に示すように特定の必須代謝産物が2つの経路から合成される場合、片方の経路を遺伝子欠失によって遮断したとしても、もう一つの経路を通して産出されるため生育に影響が表れない。しかし、このような2つの並行な経路を担う遺伝子を同時に破壊した二重遺伝子欠失株（DKO株）では致死性が現れる。このような遺伝子間の相互関係を遺伝的相互作用と呼び、細胞内機能ネットワーク解明に向けて重要な情報である。遺伝的相互作用には、図2（A）に示すように、3つの状態が考えられる。Alleviate (Buffering作用)、Neutral、そしてAggravate (合成致死)である。一遺伝子

欠失株の生育への影響が、野生型細胞の生育を1とした時に0.5へ減少する場合、AとBの二重欠失株の生育が $0.5 \times 0.5 = 0.25$ となる場合、これら遺伝子AとBはNeutral（お互いに独立）とされる。一方、AとBの二重欠失株が致死となる場合を合成致死と呼び、Aggravate (Negativeな相互作用)と言う。(B)に示す様に同一の経路の二重遺伝子欠失の場合は、一遺伝子の欠失により経路の機能が破壊されるため、それ以上の生育への影響がない。あるいは、一遺伝子欠失の影響を抑える (suppress) するような場合は、二重欠失株の生育がNeutralよりも大きくなり、Alleviate (Buffering)と言う。

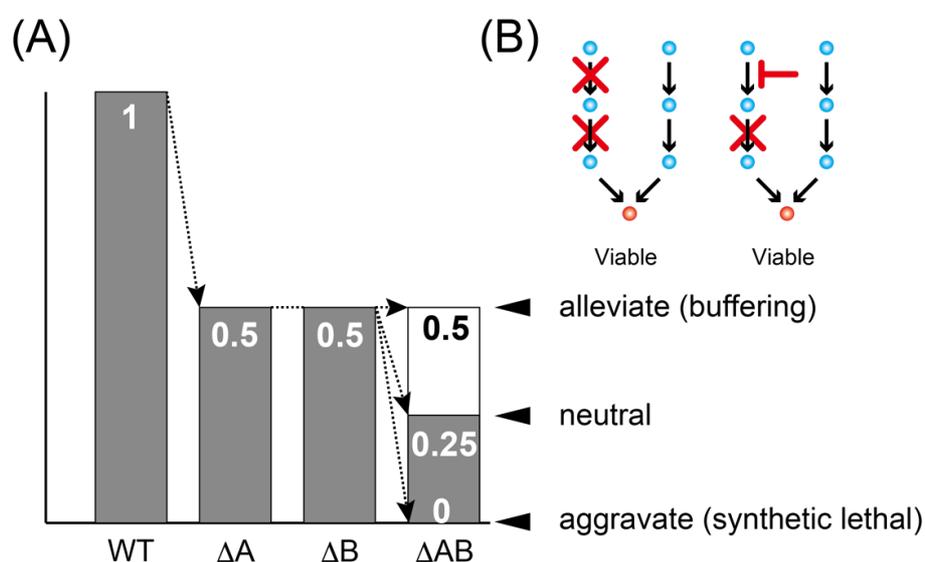


図2 遺伝的相互作用の分類

(A) 野生型細胞の生育を"1"とした場合の一遺伝子欠失株の生育および二重欠失株の生育の予想値と実測値の違いによる遺伝的相互作用の分類。(B) 同一経路における二重欠失株の遺伝的相互作用

遺伝的相互作用はこれまで機能予測やタンパク質複合体の特定等様々な研究で利用されている。例えば、酵母では本研究同様遺伝子欠失株ライブラリーを用いた網羅的な二重遺伝子欠失の生育観測実験が行われ、機能未知遺伝子の機能が推定されている(Costanzo et al., 2016)。大腸菌においても二重遺伝子欠失株の生育観測実験が行われており、代謝経路の冗長性が改めて報告された(Cote et al., 2016)。また、既知の

情報と遺伝的相互作用の情報を組み合わせることで新規のタンパク質複合体が報告された例もある(Babu et al., 2014)。このように、真核生物においても原核生物においても遺伝的相互作用を用いた機能予測や新規の発見が報告されており、未知の因子の特定において遺伝的相互作用を用いた解析は有用な方法とされている(Klobucar et al., 2018)。

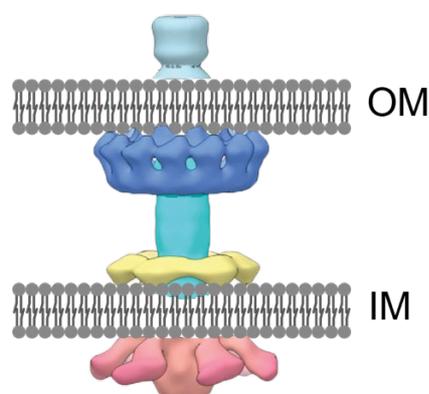
1.4 遺伝的相互作用検出のための二重遺伝子欠失実験

二重欠失株作製による遺伝的相互作用解析は、アイデアとしては単純であるが、全ての組み合わせを網羅するという事は現実的には非常に困難な課題である。2000年当時、馬場らによりKeio collectionという一遺伝子欠失株ライブラリーが構築された際は、5人のチームで丸三年をかけて構築された（森私信）。遺伝子AとBを欠失する場合、最初にどちらを欠失するかによる細胞の適応の違いから、同じ ΔAB の二重欠失株でも、Aを欠失してからBを欠失するA \rightarrow Bと、その逆のB \rightarrow Aが違う挙動を示すことは十分考えられる。それを考えると、約4,000遺伝子すべての二遺伝子の組み合わせ、すなわち $4,000 \times 4,000 = 1,600$ 万組合せの二重欠失株の作製が必要になる。後述するが、実験手法を考えても現実的に1,600万全ての組合せの二重欠失株構築を進める事が系統的な実験に適応しやすい。この事は、明らかにこれまでと同様の欠失株構築方法では現実的時間内に二重欠失株構築を終了する事は不可能である。これを解決するために、P1ファージを利用した形質導入方法(Nakahigashi et al., 2009)、もう一つは接合を用いた方法(Typas et al., 2008a)(Butland et al., 2008)、の二つの方法が検討された。最終的に、系統的な全組合せの二重欠失株構築の方向には、接合を用いた方法が選択され、改良が加えられてきている。

1.5 接合伝達を利用した二重欠失株作製手法

1946年、LederbergとTatum は、アミノ酸要求性の*E. coli*変異株を利用した実験により、*E. coli*細胞にも接合と考えられる現象を報告した(Lederberg and Tatum, 1946)。発見後多くの*E. coli*遺伝子の染色体地図が、この接合現象を用い、時間によ

る接合の中断実験などを通じ決定されてきている(Taylor and Thoman, 1964)。この接合は、本来の細胞の持つ生理現象ではなく、外来性のFプラスミドによるものであることが明らかにされている(Lederberg et al., 1952)。本来Fプラスミドが、その宿主から別の宿主へと生育範囲を広げるために移動する機構を持っていたものが、大腸菌染色体に挿入されたことで、Fプラスミド領域の移動に伴い、入り込んだ領域の染色体DNAを伴い、別の宿主へと移動していく現象である。その後の研究で、接合移動に必要な分子機構と毒素などエフェクターと言われるタンパク質群を宿主細胞へと打ち込むTypeIV分泌装置(T4SS)と呼ばれるものが共通祖先から派生していることが明らかになっている(Christie, 2001)。T4SSのモデル図を図3に示す。現在、接合によるDNA移動は、形質転換によるDNAの導入と比べて比較にならない桁違いの大きさのDNAを非常に効率よく移動させることができるため、巨大DNAの移動などの応用面でも注目されている(Zhao D et al., 2020)。現在では、クライオ電顕を用いた手法で、その詳細な構造も明らかにされており、急速に発展している分野である(Hu et al., 2019)。



Hu, B. et. al., Proc Natl Acad Sci U
S A 116, 14222-14227. より改変

図3. 接合マシンの分子構造のモデル図

私の所属する研究グループでは、Fプラスミドによる接合を利用した染色体移動に着目し、2種類の欠失株の一つの欠失を、接合移動を利用し、他方の欠失株に移動させ、その後相同組換えにより、一つの染色体上にまとめる、高効率二重遺伝子欠失株作製手法を確立させている(Typas et al., 2008a)。 Chromosomal Integration Plasmid

(CIP)は、F因子に由来する接合関連遺伝子群*tra*遺伝子群、R6Kプラスミド由来のDNA複製起点*oriRy*、及びStreptomycin(Sm)耐性遺伝子を持つ40kbのプラスミドである。染色体上特定の場所にのみ挿入できるように、ISやトランスポゾンを除き、接合および接合移動に必要な遺伝子群のみを残したプラスミドとして設計されている。CIPを相同組換えで染色体に挿入した大腸菌細胞は、CIPから供給されるTraタンパク質の働きにより接合能を獲得し、*oriT*配列に続く領域を他の細胞へと移送する。このプラスミドが染色体上に組み込まれた株はHigh-frequency recombination(Hfr)株と同様の機能を持ち、染色体を移動させる接合伝達能を有する。大腸菌株のHfr化と接合による二重欠失株作製方法を模式的に図4に示す。

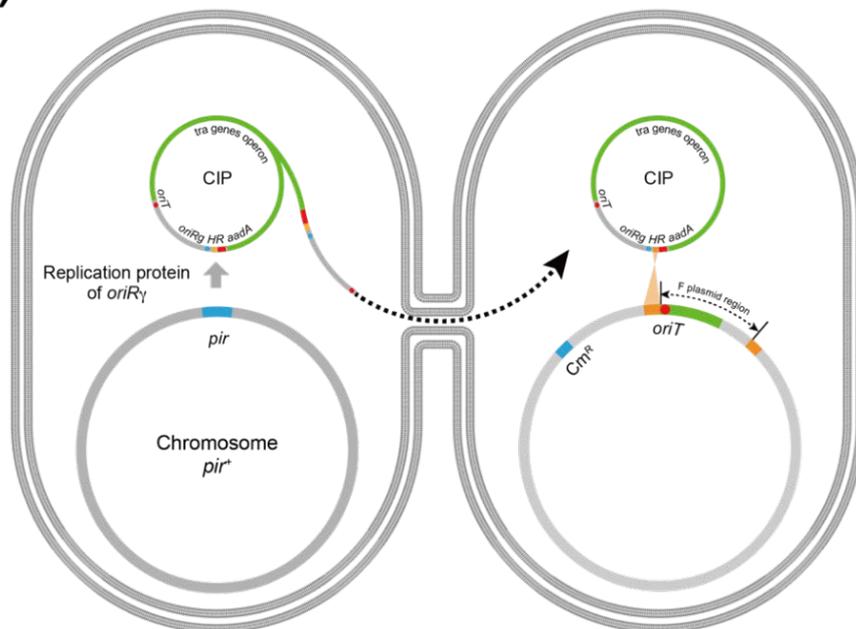
組み込まれたF plasmid配列上の接合伝達起点(origin of transfer: *oriT*)からDNAの複製に伴う接合伝達を開始され、隣接する染色体DNAを受容細胞に対し伝達することで、異なる細胞間の染色体配列の組換えを高頻度に起こす。

CIP plasmid の複製起点*oriRy*は複製に*pir*遺伝子産物を要求し、通常の大腸菌内では複製ができない。そこで、必須遺伝子である*dapA*遺伝子を破壊した大腸菌株にR6Kプラスミドの*pir*遺伝子を挿入し、plasmidの宿主とする。*dapA*遺伝子の欠失株はジアミノピメリン酸(DAPA)を添加した培地で培養することで生育が可能である。この宿主を任意の受容菌と共存させることで、CIP供与菌と受容菌の間の接合伝達によりCIPが受容菌に移動する(図4A)。受容菌は*pir*が存在しないため、CIPをプラスミドとして保持出来ないためCIPが相同組換えで染色体に挿入されHfr株となる。また、DAPA非存在下においてストレプトマイシンで選択することで、相同組換えによりCIPが染色体ゲノムに組み込まれSm耐性遺伝子を獲得した受容菌のみ生育する事が可能になり、簡便にHfr化された株のみを選択する事ができる。これにより*dapA*変異を持たないKeio collectionとは違う薬剤耐性遺伝子を持つ新たな欠失株のHfr化を行う。

Hfr化した欠失株(ASKA barcode deletion collection、Cm耐性)を供与株とし、寒天培地上に高密度で増殖させたKeio collection の一遺伝子欠失株のコロニーと接触させることで、約4000株のKeio株にASKA株側の遺伝子欠失領域を移動させることができる。ASKA株は遺伝子欠失領域にCm耐性遺伝子を有し、Keio株はカナマイシン(Km)耐性遺伝子を有するため、生育したコロニーをKmとCmの両方を添加した培地

で培養することで、ASKA株の遺伝子欠失領域が相同組換えによってゲノムに取り込まれた二重欠失株を選択することが可能である（図4B）。

(A)



(B)

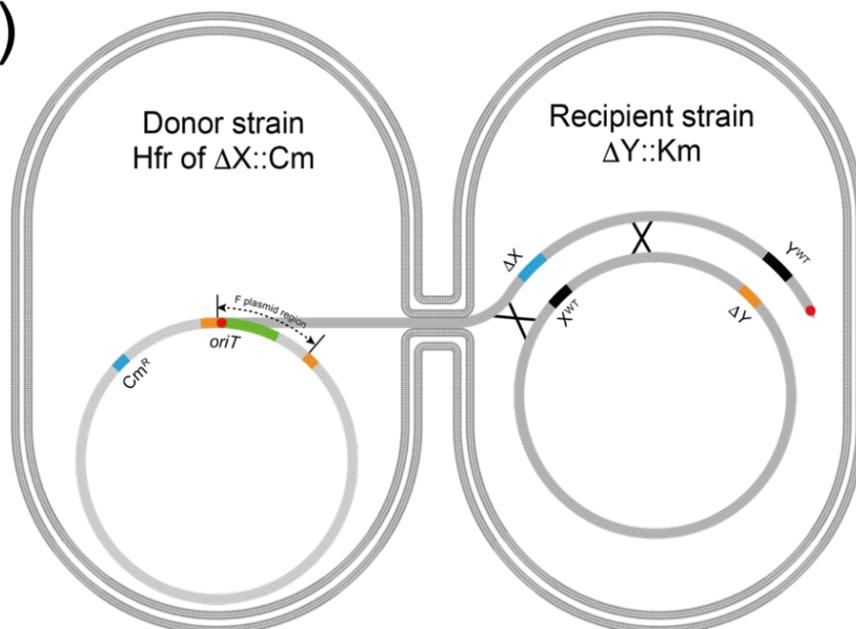


図4. CIPによるHfr化と接合による二重欠失株作製方法

(A) CIPによる大腸菌のHfr化。CIPは完全長の*tra*遺伝子群（緑）を持ち、その他のISやトランスポゾンを持たない。複製起点も本来の*ori2*は削除しており、代わりに*oriRg*で複製を行う。CIPは*pir*を持つ大腸菌内で複製可能であり、供与菌として機能する。Fプラスミドを持たない受容菌と出会うと接合を開始し、CIPが受容菌側へ移る。受容菌には*pir*が存在しないので、CIPは複製されず、CIPの持つSm耐性で選択すると、CIPが染色体へ染色

体上の相同領域との間で相同組換えで挿入されたもののみが取得できる。(B) Aの手法でHfr化した欠失株は、Hfr株として機能し、Fプラスミドを持たないKeio株が存在すると、接合が始まり、供与菌側のoriTから染色体移動が始まり、供与菌側の欠失を受容菌側に移動させ、KmとCmの2重の選択で、二重欠失株が構築される。

1.6 Colony-live について

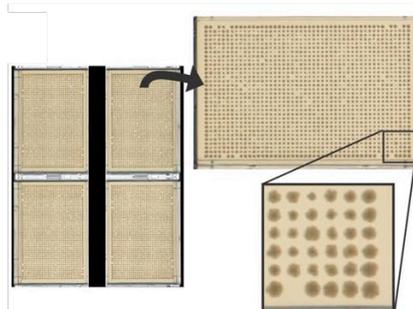
1990年代のゲノムプロジェクトによって、シーケンサーをはじめとした大規模な自動化装置の開発が急速に押し進められ、1997年には大腸菌の全ゲノム構造が明らかになったことにより、複雑な生物学的プロセスに関与するすべての遺伝子産物を単一の生物の中で同定することが原理的に可能になった(Blattner et al., 1997; Yamamoto et al., 1997)。また、系統的に作製された全遺伝子を網羅する実験リソースの開発などにより(Kitagawa et al., 2006) (Baba et al., 2006)、大規模な発現解析(Oshima et al., 2002a; Oshima et al., 2002b)やタンパク質相互作用解析(Arifuzzaman et al., 2006)など、いわゆるオミックス研究が可能にされた。その中で微生物のコロニーに特化した高効率かつ菌株のスポット精度が高い細胞植菌ロボットがトロント大学とSinger社によって開発され、遺伝子欠失の表現型の大規模な観測と解析に使用されている。当初はカメラを用いた一定時間後の画像による、その生育量を定量するものであった(Typas et al., 2008b) (Butland et al., 2008)。私自身も系統的解析手法の開発に携わっており、これまでにColony-liveというシステムをメンバーの一員として共同構築した(Takeuchi et al., 2014)。Colony-liveとは、経時的観察による生育の変動と細胞massの定量を目的として、透過型スキャナーを用いたコロニー定量装置である。市販のスキャナーを利用して経時的に取得した画像を、画像解析技術により各コロニーを定量するソフトウェア群によって、高効率かつ高精度な大腸菌のコロニーの生育観察を可能にしたシステムとなっている。また、寒天培地の厚さや細胞の植菌量、コロニーの位置による生育の違いによるバイアスやプレート間の生育状態の違いの補正やコロニー間に働く隣接効果といった遺伝子欠失以外の要因を除く処理も酵母で開発された手法をシステムに組み込んでおり(Baryshnikova et al., 2010)、既存の網羅的な生育観測の実験系に比べてクオリティの高い観測が可能となっている。現在は、1台のスキャナーにつき4枚のプレートのスキャン画像を取得し、そのプレート画像の1枚ずつの切り出し、切り出した画像から寒天培地上のコロニー画像の取得及びコロニーの同定と、透過型スキャナー

による擬似的な細胞mass量の定量と補正・標準化を行うシステムとして公開している。

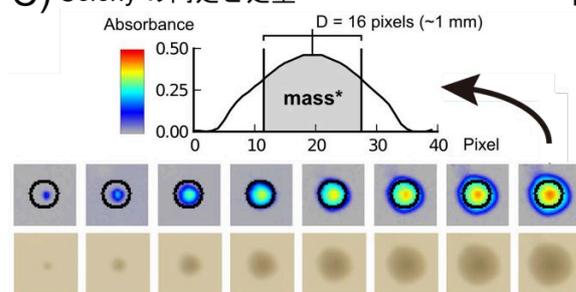
A) scannerによる自動画像取得



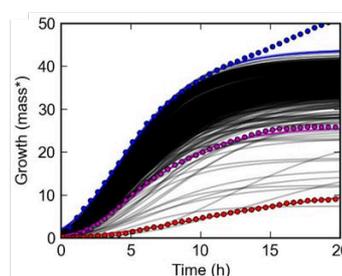
B) 取得画像の切り出し



C) Colonyの同定と定量



D) モデル式へのフィッティングと定量



Takeuchi, R., et al., BMC Microbiol 14, 171 より改変

図5. Colony-liveシステム

A)当研究室では2台のPC-Linuxマシンで12台の透過型スキャナーを制御し、一定時間毎に自動的に寒天培地上のコロニーの画像を取得する。B)全観測を終えると、自動的に各プレートの画像からプレート毎の画像を切り出しを行う。C)その後最後の画像を用い、各コロニーの中心位置の同定を行い、中心から直径16ピクセル（EPSON GT970の画素数の場合）内のすべてのピクセルのスキャンデータを取得し、細胞massとする。D)各コロニーのmass量をモデル式にフィッティングを行い、定量を行う。これらは、プレート間の補正などもシステムに組み込み、自動で処理を行う。得られたデータは、サーバー上のデータベースに自動的に登録され、解析に供される(Takeuchi et al., 2014)。

システム自体の改良は現在も続いており、自動化をさらに進め利便性と信頼性を高めてきている(図6)。さらに大きな改善点としてこれまでのスキャンソフトを介したスキャナーの制御から、SANE(<http://www.sane-project.org/>)を用いて直接スキャナーの制御を行うシステムへの変更を進めており、一部本研究においてもテストを行なっている(野村、私信)。現在は、12台のスキャナーを2台のコンピューターで制御しているが、新しいシステムでは、1台のコンピューターですべてのスキャナーが制御可能となる。画像取得の流れと定量化・解析においての変更はない。

Colony-live system

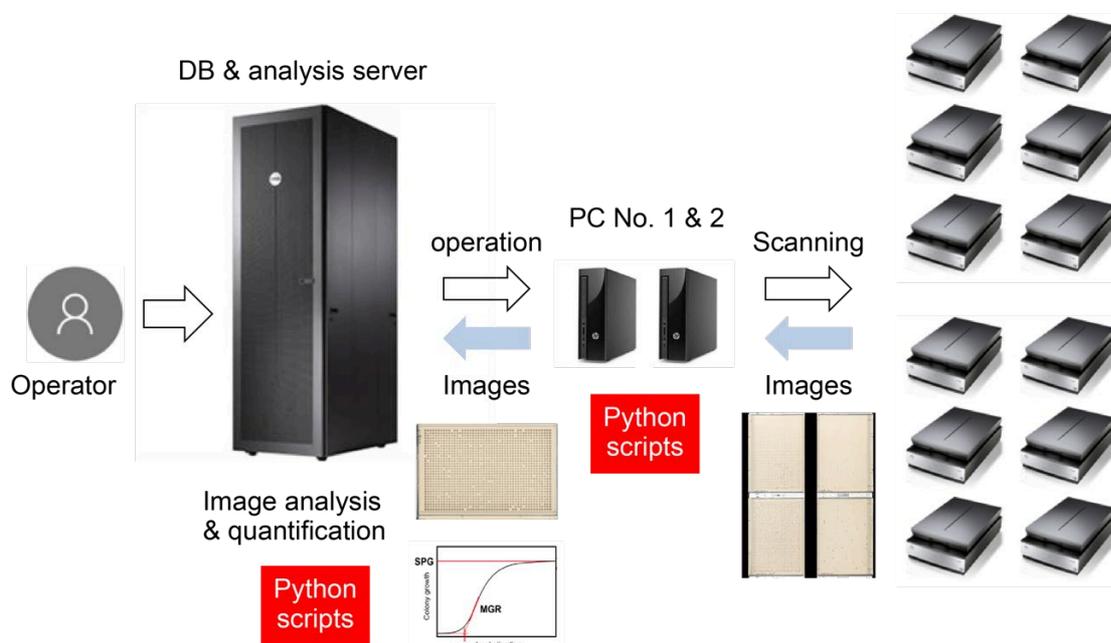


図6. Colony-liveコロニー定量・解析システム

12台のスキャナーを2台のPC Linuxで制御し、画像を一定時間ごとに取得する。取得された画像は一時的に制御コンピュータ上に全観測が終了するまで保管される。画像取得後はプレート画像の切り出し、コロニー同定、コロニー領域の透過スキャンングによる定量値を取得後、モデル式へのフィッティングを行うことで、各種生育パラメーターの取得が可能である。最終的なデータは、全てサーバーのデータベースシステムに登録される。

実験者はWEBを通して実験情報の入力を行い、画像取得を開始させるのみで観測が可能である。設定した時間まで画像取得を行なった後、取得された画像は全てDB & 解析サーバー内に移され、Colony-liveの定量プログラムによって画像のコロニーは全て定量化されデータベースに時系列順に登録される。時系列に並べられたコロニーの定量値はコロニーの面積を表すCONVの計算と、Gompertz関数式(Benjamin Gompertz., 1825)へ回帰してLag time growth(LTG; 生育開始までの時間)、Maximum growth rate(MGR; 最大生育速度)、Saturation point of growth(SPG; 最大到達濁度)の3つの生育度スコアを計算し、産出した生育スコアはデータベースに記入する。データベースはリレーショナルデータベースのMariaDB(MariaDB_foundation)に登録する。CONVは、カメラによるコロニー画像取得方法と同様のコロニー底面積の定量値である。以上の

動作は全て自動で行われるため、使用者は実験条件と撮影時間の登録を行うだけで観測したいプレートのコロニーの生育が標準化されたスコアを得る事が出来る。

1.7 未知の代謝経路がもたらす代謝シミュレーションの偽陰性の解明に向けたアプローチ

1980年代にヒトゲノムプロジェクトの必要性が議論されだし、1986年にはDulbeccoがガン研究に端を発したゲノム解析の必要性を説いた(Dulbecco, 1986)。多くの議論がイギリス・アメリカを中心に進められ、最終的に1990年に始まったゲノム研究と時を同じくして、ゲノムレベルの代謝経路情報の数学的モデル化が、24遺伝子、14酵素反応で始められた(Majewski and Domach, 1990)。その後Palssonのグループが大腸菌の大規模な知識集積とモデル化を始め、1993年には53酵素反応、30代謝産物のモデル化を行い(Varma A., 1993)、大腸菌代謝機能のシミュレーション実験が可能となった。現在ではMycoplasmaにおいて全遺伝子モデルが構築され、シミュレーションが可能となっている(Karr et al., 2012)(Kaizu Kazunari., 2020)。代謝ネットワークの再構築とは、生化学的、遺伝学的、ゲノム学的に構造化された生物に関する詳細な情報を、構造化されたネットワークの形式で格納したものである(Reed et al., 2003)。形式とは、例えばネットワークのノードが代謝化合物、エッジが代謝反応で結ばれたグラフとなっており、どのノードにどのエッジが引かれているか、そしてどれだけの速度で酵素反応が進むのかといった情報が記載されている。特に大腸菌の代謝ネットワーク再構築は、ゲノムスケールにまで到達し、代謝工学、モデル指向の発見、表現型スクリーンの解釈、ネットワーク特性の解析、進化過程の研究の5つの主要なカテゴリーにおいて、基礎的・実用的な応用に幅広く活用されている。代謝工学研究においては、代謝ネットワークモデルを用いた*in silico*株の設計を行い所望の生成物を過剰生産する株の候補を設計する際の判断材料としてシミュレーション実験が行われている(Fowler et al, 2009)。モデル指向の発見においては、代謝ネットワークに含まれる現在の生化学的・遺伝的情報に加えてある機能に関連したタンパク質配列や遺伝子を持たない酵素反応を担う遺伝子の予測など、生物学的発見を促進するために役立てられている(Fuhrer et al, 2007)。また、生物学的ネットワークの本質的な特性を解析・解釈するために利用されており、この一例として異なる生育条

件での連成反応活性の発見が挙げられる(Anthony P et al, 2004)。進化研究ではモデルは容易にネットワーク再構成が出来る点から、適応進化イベント、水平遺伝子導入、最少代謝ネットワーク進化における細胞ネットワークの検討に使用されている(Pal et al, 2005, 2006)。

また、同時期に代謝モデルを使用したシミュレーションのアルゴリズムの開発も進められた。その中で現在最も広く使用されているものがFlux balance analysis(FBA)(Orth et al., 2010)である。FBAとは代謝モデルを解析する数学的手法であり、代謝ネットワークモデルを用い、代謝反応によってどれだけの量を流す事が出来るかの範囲、つまり代謝反応の流束に対して制約をかけその範囲内で収まるように目的関数としてMGRを定義し出力である細胞のバイオマスが最大となるような解を求める手法である(Edwards J., 2001)。また、大腸菌の代謝ネットワークのConstraint-based modelを使う事で培地条件を変更するために入力となる基質に制約を付けたり、エッジに更なる制約を付ける事で遺伝子欠失を再現することが出来、生成物の自由エネルギーの方が反応物質の自由エネルギーより低いという熱力学原理にもとづいて化学的反応の方向性に制約を付ける事で使用者の望んだ実験条件を作り出す事が可能である。これらの制約条件はモデル内で定量的に記載されており、数値を変更する事で実験条件の変更が可能となっている。デフォルトのパラメーターは、データベースの情報を元に数値化した値がそれぞれ設定されている。FBAは入力と総バイオマスの量が等しいという定常状態を仮定、つまり入力と出力が等しいと仮定しノード間の反応を方程式で表現する。FBAがどのようにシミュレーションを行うのかより詳しく解説する。

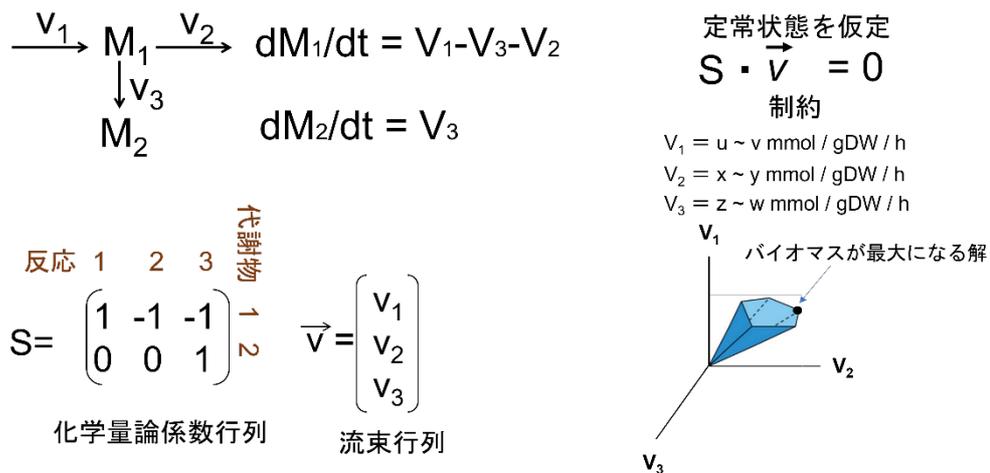


図7. FBAの概略

代謝ネットワークを流路と考え、inputとoutputは同じ量であることを利用してモデル化を行う。その際、定常状態を考えると、 $dS/dt = 0$ となり計算を簡略化できる。

まず、代謝ネットワークは図7のように代謝物質の濃度変化を微分方程式として表現することができる。ここで、定常状態であるとするこれら3つの微分方程式の値は0となる。定常状態を仮定することで微分方程式を解く必要が無くなったため計算量を非常に小さくでき、高速なシミュレーション実験が可能となる。また、方程式は化学量論係数行列(stoichiometric matrix, S)で表すことが出来るようになり、各次元が化学量論係数行列に対応する反応の流束率を表わすベクトル \vec{v} と掛け合わせたものは $S\vec{v} = 0$ となる。この関係を満たすベクトル \vec{v} の解空間は先ほど述べた様々な制約によって狭められ、その中から出力されるバイオマスの量が最大となるような最適解が選ばれる。こうして得られた解から、使用者はどの生成物が指定した制約のもとでどれだけ産出されるのかという情報や細胞の生育度の予測値を得る事が出来る。細胞内では実際に酵素が存在するが反応が分かっていない等、まだまだ生物的知見の不足から厳密な計算が難しい状況ではあるが、現実的レベルで正確な解を得ることができている(Monk et al., 2017)。FBAで最も使用されている大腸菌の代謝ネットワークモデルはPalssonらが構築した*Escherichia coli str. K-12 substr. MG1655*株の代謝ネットワークモデルである(Orth et al., 2011)。2003年に最初に公開されたiJR904には904個の代謝に関わる遺伝子とそれに関係する代謝産物と代謝反応が記載されていた(Reed et al., 2003)。

このモデルに代謝に関わる要素を加えるアップデートが幾度も行われ、2007年に iAF1260(Feist et al., 2007)、2011年に iJO1366 (Orth et al., 2011) 、2017年に iML1515 (Monk et al., 2017) が発表された。公開されているFBAを行うPythonプログラムである Cobrapy (Ali Ebrahim et al., 2013)は代謝ネットワークモデルには今のところiJO1366が設定されている。

2012年、データベースに蓄積された一遺伝子欠失株の生育実測値と、iJO1366とFBAを用いて同条件下で行われたシミュレーション実験で得られた生育の予測値との比較が行われ、実測と予測の間には乖離（ギャップ）が存在することが報告された(Orth and Palsson, 2012)。そのうち、実測値は生育を示すが予測値は致死を示す偽陰性(False negative: FN)の場合に関しては、未知の酵素または経路が頑強性に寄与している事が原因であると言及された。

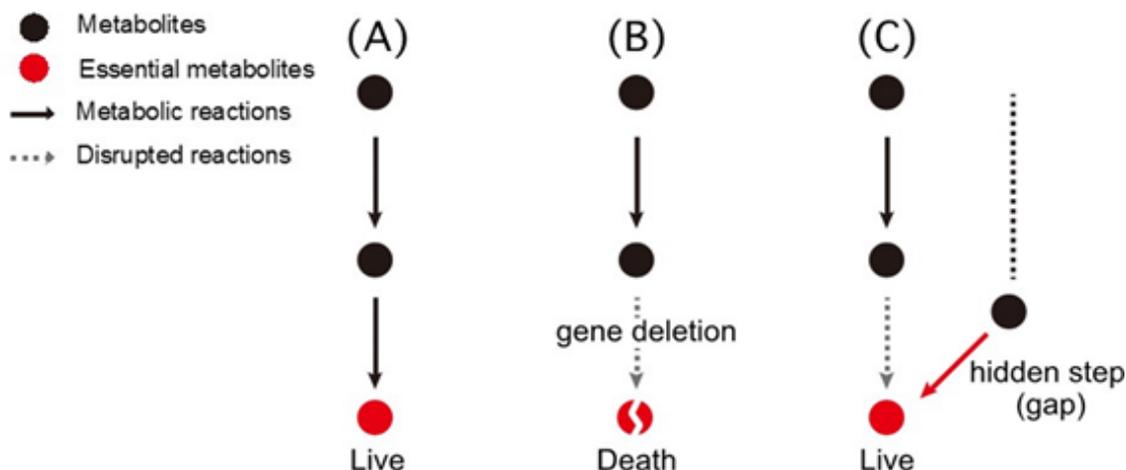


図8. 偽陰性を示す機構

黒丸は代謝産物、赤丸は必須代謝産物、実線の矢印は酵素反応、破線の矢印は遺伝子欠失で破壊された酵素反応ステップを示す。(A)では必須代謝産物を正常に生産できるので細胞は生育する。(B) 必須代謝産物合成経路の一つのステップが遺伝子欠失で酵素が供給されず破壊されるため死に至る。(C) 一方単一の経路と考えられていたステップを破壊しても生存している場合、隠された経路の存在が考えられる。これが偽陰性を生じる原因の

1つである

上の図8のAは必須の代謝産物が1本の経路から生成されている様子を表している。この時、図8Bのようにこの経路が遺伝子欠失によって遮断された場合、必須の代謝産物が生成出来なくなるため致死となる。しかし、図8Cのように未知の代替経路が存在するならば生育は可能となる。この時、代謝ネットワークモデルは既存の経路しか使えないが実際は代替経路が存在するため実験との間に差異を引き起こす。これがFNの原因であり、FNを示す遺伝子を対象としたDKO株生育観測実験を行い遺伝的相互作用を示す遺伝子を検出することで、頑強性の要因となっている未知の因子を見つけることが出来る可能性を示唆している。当研究室では、前述の通り接合と **Keio collection** 及び **ASKA barcode deletion collection** を組み合わせた網羅的なDKO株作製手法と、コロニーの生育度を網羅的に測定するシステムである **Colony-live** の開発を行った。

本研究では、大腸菌代謝ネットワーク構造の完成を目指し、シミュレーション結果と一遺伝子欠失株生育の実測の比較から得られたFNを示す遺伝子情報を元にネットワーク上のギャップを探索し、そのギャップを埋める酵素および経路を同定することを目的とする。FNを示す遺伝子をターゲットとし、**Keio collection** と組み合わせて網羅的な二重遺伝子欠失株を作製し **Colony-live** を用いた解析から未知の代謝経路の探索を行う。現段階において、大腸菌にはこのような網羅的なDKO株の定量データは存在せず、代謝経路における頑強性の解析において多大な貢献を与えることが出来ると思われる。

2 章

大腸菌一遺伝子欠失株の実測とシミュレーション実験との間に生じるギャップ

2.1 緒言

2.1.1 最少培地で偽陰性 (False negative: FN)を示す遺伝子の特定

偽陰性 (False negative: FN)を示す遺伝子については序論で述べたように、その遺伝子を欠失した条件において、モデルには存在しない隠れた酵素あるいは代替経路の存在が強く示唆される。モデルと実験の食い違いの原因を調べる事は、この隠れた経路の存在を探索することであり、細胞内機能ネットワークの完全な理解につながるものである。FNを示す遺伝子特定のための代謝経路を対象とした実験を行うにはLBのような栄養が豊富な培地では特に代謝関連遺伝子の単一遺伝子欠失の表現型変化を観察することは難しいと考えられる。そのため、M9最少培地に種々の炭素源を添加した培地を使用し、生育実験およびシミュレーションを行い、得られた定量値の比較を行う事でFNの特定を行う。最少培地に関して、よく利用されるいくつかの培地があるが、MOPS培地やME培地を利用しない理由は、PalssonらによるiJO1366の構築時にM9での実験データが利用されてきたためである。一般的に利用されるM9培地に、微量元素群を補填する方法も、Palssonらの方法に従い、条件を揃えている。本研究ではM9に炭素源としてGlucose、Glycerol、Succinate、Xyloseの4種類の糖源をそれぞれ培地に添加し、網羅的な一遺伝子欠失実験を行う。一遺伝子欠失株には、当研究室が保有するKeio collectionを使用した。また、これまでの観察実験では、Keio collectionが保管されているGlycerol stockから植菌する段階においてLB寒天培地を使用していたが、LB寒天培地に含まれる栄養の持ち込みにより株の生育度に対して影響を与える場合があることが報告されている(Orth and Palsson, 2012)。これは、大腸菌にとって必須の代謝産物であるが必要な量が微量であるものが、LB培地からM9最少培地への植菌の際にわずかに持ち込まれた結果、生育に影響が出てしまう事が原因と考えられる。このため、本来M9培地では働く合成経路が残留しているLB培地の栄養を使用することで働かなくなる可能性が考えられる。使用される代謝経路のネットワークが変化してしまうと、M9培地を生育条件として設定した代謝シミュレーション実験と比較するには不適切なデータとなってしまう。そのため、今回の一遺伝子欠失実験で使用する培地はM9に糖源を加えた寒天培地に統一し、網羅的な生育観察実験を行い、シミュレ

ーション実験の結果と比較する事でFNを示す遺伝子の同定を試みた。

2.1.2 最少培地でFNを示す遺伝子を対象とした二重欠失株(Double KnockOut: DKO)解析において発生した問題

序論で述べた通りFNを示す遺伝子群は、隠された代替酵素・経路あるいは補償回路の存在を強く示唆するものである。当研究室では既に接合を用いた高効率なDKO株作製手法が構築され、栄養培地のLB培地を用いたDKO株生育観測実験の実績がある。そこで、FNを示した遺伝子を対象に最少培地であるM9 Glucose培地を利用した網羅的DKO株構築による大規模遺伝的相互作用解析により、隠された代替酵素・経路及び補償回路を構築する候補遺伝子群の同定を行うこととした。

二重遺伝子欠失株の作製のための接合実験において、受容菌(Keio collectionの欠失株)が持つカナマイシン耐性と供与菌(ASKA barcode collectionの欠失株)が持つクロラムフェニコール耐性の両薬剤による選択の段階で、殆どの株が生育できず観測が不可能な状態であった。原因として、培地の違いによる接合での二重欠失株構築の効率の違いが疑われた。そこで最少培地での接合による二重欠失株構築方法の確立を行った。接合過程は、図9に示す様に、1) DNA移動開始段階、2) 受容菌との間の接合機構確立段階、3) 受容菌内へ移動後の相同組換え段階の3つに分けて考える事ができる。培地の違いが、どの段階に大きく影響するか明確ではないため、それぞれの段階での個々の事象の培地による効率の違いを定量化することが必要であるが、まずは現実的解決法として、二重欠失株取得の効率を指標として、手法の改善に取り組んだ。

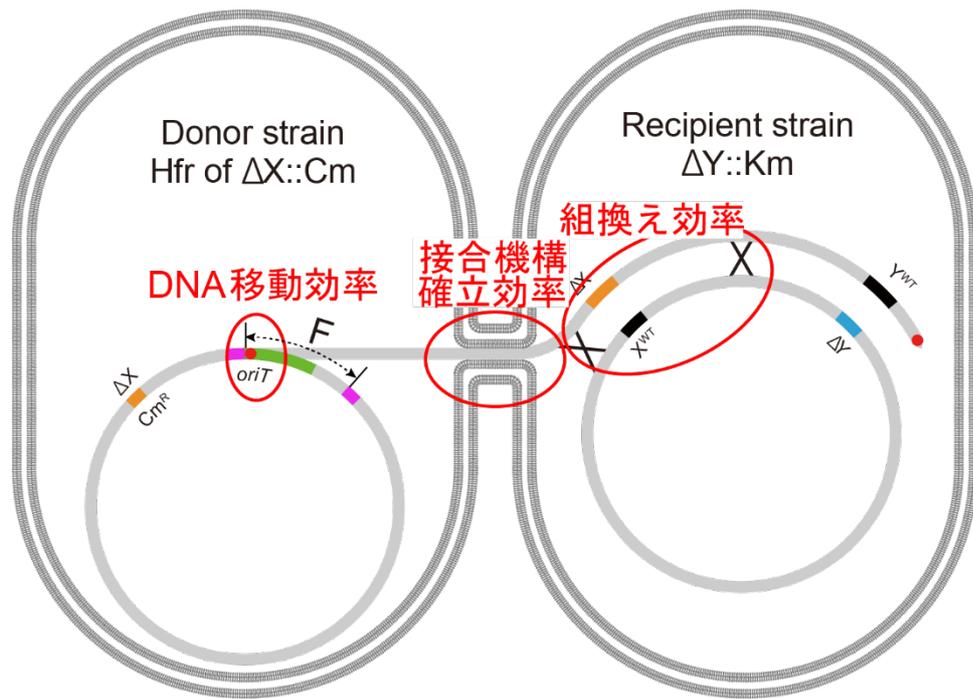


図9 接合機構における3つの効率
 接合伝達による組み替え体の構築は、1) DNA移動開始段階、2) 受容菌との間の接合機構確立段階、3) 受容菌内へ移動後の相同組換え段階の3つの段階からなる。

2.2 実験条件

2.2.1 培地及び化学物質

培地は、基本的にMolecular Cloning 4th Edition(K Roberts., 2015)に従った。M9培地に関しては、Palssonらの方法に従い微量金属元素の補填をTrace element(表3)を加えることで行った(Monk, 私信)。

表3. Trace element(100ml)の組成

化学物質	量
FeCl ₃ · 6H ₂ O	0.7g
ZnCl ₂ · 4H ₂ O	0.2g
CoCl ₂ · 6H ₂ O	0.2g
Na ₂ MoO ₄ · 2H ₂ O	0.2g
CaCl ₂ · 2H ₂ O	0.1g
CuCl ₂ · 6H ₂ O	0.13g
H ₃ BO ₃	0.05g
ddH ₂ O	Up to 100ml

オートクレーブ滅菌し、使用時にM9培地500mLに対して、Trace element を125 μ L入れ、調整した。

Glucose, Glycerol, Succinate及びXyloseは20%溶液を作製し、終濃度0.4%になるように使用時に添加し使用した。寒天培地には1.5%になるように寒天を加え、使用した。化学物質は基本的に和研薬から購入した。

薬剤及びDiaminopimelic acid (DAPA)は、表4に示す最終濃度で使用した。

表4. 薬剤及びDAPAの濃度

薬剤	濃度
クロラムフェニコール	30 μ g/mL
カナマイシン	25 μ g/mL
ストレプトマイシン	15 μ g/mL
DAPA	0.3mM

2.2.2 菌株及びプラスミド

SKO株ライブラリーにはカナマイシン耐性をもつKeio collection(Baba et al., 2006)及びクロラムフェニコール耐性を持つASKA barcode deletion collection(Otsuka et al., 2015)を使用した。なお、Keio collectionは独立の2クローンが存在しているが、Partial duplicationのような問題が確認されていない1クローンを選別し冗長性を除き、並べ替えを行ったライブラリーを使用した。この並べ替えを行ったライブラリーは、384プレートで保管しているが、最終的にプレートあたり1536の密度にコロニーを形成させると、1536プレートの最外周には野生型の大腸菌が並ぶように配置を行っている。さらに、1536の密度であると、大腸菌全遺伝子は3枚のプレートで網羅できる。3枚のプレートを区別できるように、最外周の一部のコロニーを除くことで、最外周のコロニーの生育パターンでプレートを区別できるようにしている。

1536プレートの作製手順は以下の通りである。まず、384マイクロタイターにグリセロールストックとして保管している並べ替えを行ったKeio collection大腸菌ライブラリー菌液を室温にて溶解後、Singer RoToR HDA (Singer Instruments, Somerset, United Kingdom)を使用して、Singer RoToR用のplateに作製しているそれぞれの糖源とKmを加えたM9寒天培地に384のLong pin padを用いてスタンプを行い、37°Cで一晩培養を行う。翌日、384密度のコロニーを384のshort pin padを用い、4プレートを1枚に集積することで、プレートあたり1536コロニーの高密度コロニープレートが出来る。SKO株の生育観測にはこの1536プレートを行い実験に使用した。

DKO実験における供与菌には、ASKA barcode deletion collectionから欠失株をCIPプラスミド[竹内博士論文]によってHfr化したものを利用した。接合実験の受容菌には、並べ替えを行ったKeio collectionを用いた。

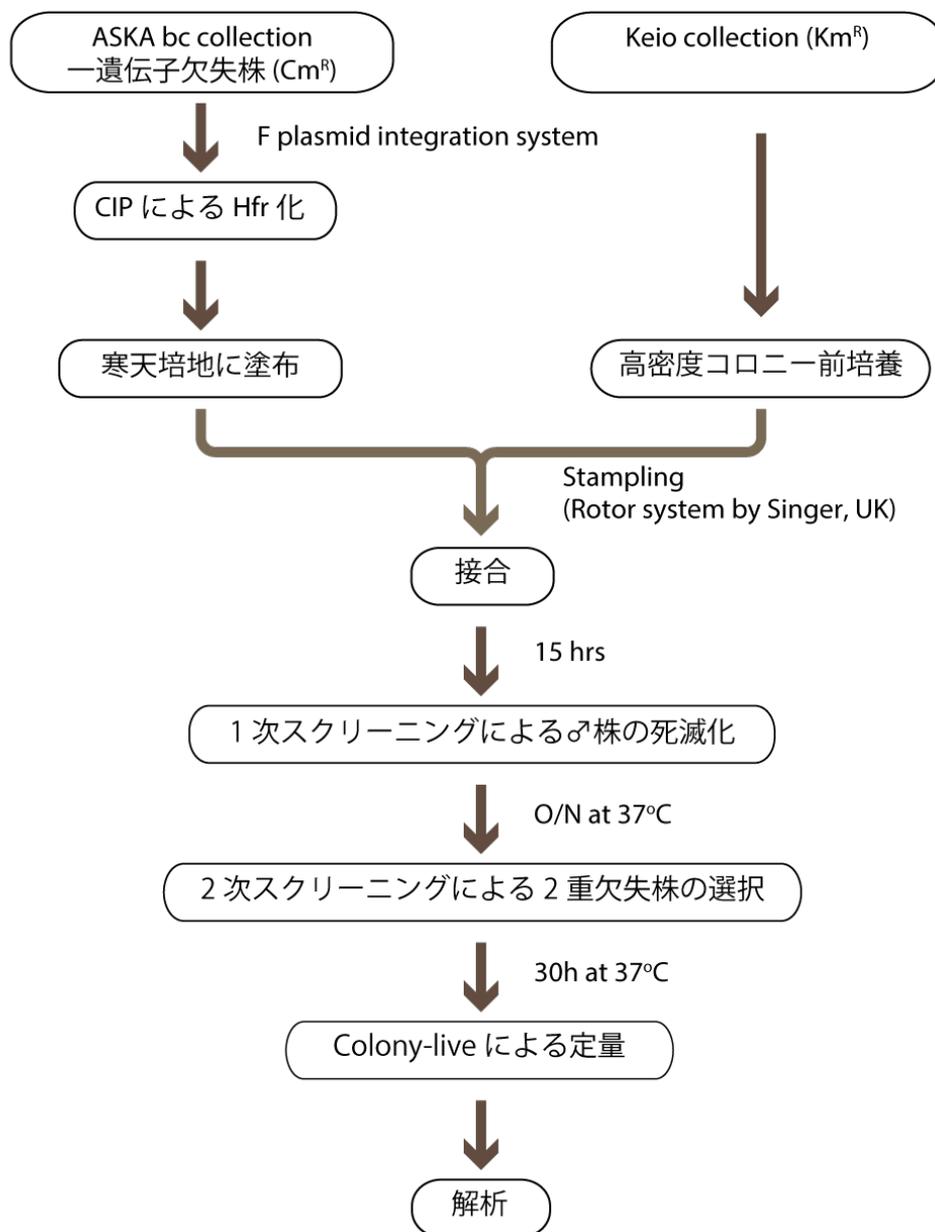
2.2.3 一遺伝子欠失株および二重欠失株コロニーの生育観察

並べ替えを行った1536フォーマットのSKO株について、Colony-liveを用いて観測実験を行い生育度を得た。観測は30分間隔に設定し、30時間培養の自動計測を行った

二重欠失株の観測実験では、受容菌として並べ替えを行った1536フォーマットのKeio collection(図10. E)を使用した。供与菌として、FNを示す遺伝子をASKA barcode deletion collection (図10. A)を、CIPプラスミドを用いて染色体上に挿入しHfr化したも

の(図10. C)を使用した。CIPプラスミドはASKA barcode deletionの欠失遺伝子領域のゲノム上の位置が3628992bpから4641652bp又は0bpから1775220bpまでの場合はCIP8を、1775221bpから3628991bpまではCIP16を選択し使用した。DKO株はHfr化した供与菌を塗布した薬剤無しのM9 Glucose寒天培地(図10. D)へ、Singer社RoToR用プラスチック1536 short pin padを用いて1536フォーマットのKeio collectionをスタンプして植菌し、一晚培養することで接合(図10. F)を行い、Hfr株の染色体領域を受容菌に伝達させ作製した。Keio株とASKA barcode deletion collection 株は、共通祖先の大腸菌から構築しており、ほとんどの染色体領域が同じである為、染色体相同領域間で組換えが起こり、接合により運ばれてきた供与菌側の欠失領域が受容菌染色体に挿入され、二重遺伝子欠失株が出来る。接合後のコロニーを、Kmが入ったM9寒天培地に1536 short pin padを用いてスタンプして植菌し、一晚培養で選択を行う事で供与菌を除去した(図10. G)。さらにKm及びCmを入れたM9寒天培地に1536 short pin padを用いてスタンプし受容菌を除去し組換え株である二重欠失株を選別する(図10. H)。最終的にDKO株のみとなったプレートを恒温器中に設置したスキャナー内部に並べ、測定を開始し

た(図10. I & J)。



2重欠失株作製方による遺伝的相互作用解析の流れ

図10. DKO実験の流れ

A)~J) LB培地を用いたDKO実験のために構築されたプロトコル

2.2.4 シミュレーション実験に使用したソフトウェアと条件

代謝ネットワークによる一遺伝子欠失シミュレーションには、iJO1366のモデルを使用した。実際の計算には、openCOBRA(<https://opencobra.github.io>)のPythonモジュールであるCOBRAPy (Ebrahim et al., 2013)を用いた。言語はpython2.7.10を使用し、COBRAPyのバージョン0.3.2をインポートしFlux Balance analysisシミュレーションを行った。COBRAPyは標準設定で培地条件がM9 Glucoseとなるように設定されており、Glucoseの取り込みの流束は-10mmol / gDW / hになっている。そのため、栄養源としてM9 Glucose以外の3つの炭素源を設定する際には、Glucoseの取り込みの流束の値を0にし、次に使用したい炭素源の流量の値のみを変更することで、今回行った大腸菌の一遺伝子欠失実験と同様の条件でシミュレーションを行うことができる。COBRAPy 0.3.2では

1) iJO1366モデルを選択

```
model = cobra.test.create_test_model("ecoli")
```

2) 糖源を変更するため、グルコース取込み流速を0にする。

```
model.reactions.get_by_id("EX_glc_e").lower_bound = 0
```

とすることでGlucoseの取り込みの流束の値を0にする事が出来る。"EX_glc_e"はiJO1366内でGlucoseの取り込みの流束に紐づけられているIDであり、それを"0"にすることで、Glucoseの供給を計算上から削除する。

今回使用するそれぞれの炭素源の流量の値は標準設定のM9 Glucoseの流束に合わせて全て-10mmol / gDW / hとし、一遺伝子欠失株の網羅的な生育シミュレーション実験を行い、それぞれの遺伝子欠失が示す細胞の生育予測値を計算した。例えば、M9

Glycerol上での一遺伝子欠失シミュレーションは以下の通りコマンドラインに入力することで行う事が出来る。

```
import cobra.test
from cobra.flux_analysis import single_gene_deletion
model = cobra.test.create_test_model("ecoli")
1) glucoseを環境中から削除
model.reactions.get_by_id("EX_glc_e").lower_bound = 0
2) glycerol取込み速度を10に設定。
model.reactions.get_by_id("EX_gly_e").lower_bound = -10
3) 一遺伝子欠失株を用い、生育の計算を行う。
growth_rates, statuses = single_gene_deletion(model)
```

以上の設定により、モデルとしてiJO1366を用いた一遺伝子欠失株による網羅的コンピュータ生育実験を行った。

ハードウェアは、IBM社製ノートPCで行った。また、使用したColony-liveはHP社製デスクトップ型PC 2台にインストールしEPSON社整透過型スキャナーGTX 970 6台ずつ合計12台を接続し使用した。データベース及び定量解析は、PC Linux server (28CPU, 256GB)を使用し、解析はPython2.7.10を、データベースはMariaDB version 5.5.56を使用した。

2.2.5 最少培地上でのDKO株作成プロトコルの見直し

最少培地を用いたDKO株作製は、これまで当研究室で栄養培地において使用されてきたプロトコルでは殆どの場合定量化を行う事ができなかった。このため、接合による二重欠失株作製方法の改善が必要となり、

- 1) 前培養方法
- 2) 接合時間及び接合時の培地環境

3) 接合移動後の供与菌、受容菌、組換え後の二重欠失株の選別方法
の各項目の検討を行なった。

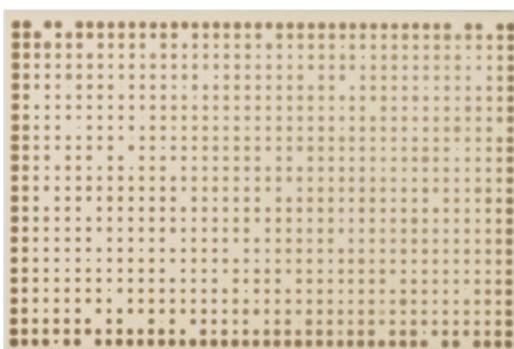
方法は、培地環境、時間、選別方法であり、培地等はMolecular Cloning 第4版に従った。

2.3 結果

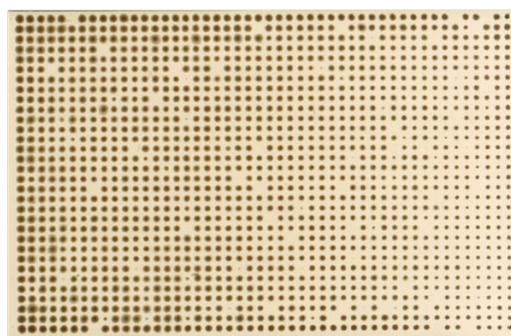
2.3.1 糖源の違いによる実測値

4種類の糖(Glucose、Glycerol、Succinate、Xylose)をそれぞれ唯一の炭素源とした条件でのシミュレーション実験と生育実験の食い違いを解析することを目的として、一遺伝子欠失株ライブラリーのKeio collectionをM9寒天培地に各糖を炭素源として添加した培地上に1536の密度でスタンプし、生育をColony-liveで解析し比較を行った。観測実験は3回繰り返し、M9 Glucose、M9 Glycerol、M9 Succinateに関しては全ての実験で定量化を行うことが出来た。しかし、M9 Xyloseに関しては図11から分かるように、明らかに他の糖源と比較して生育が悪く、定量可能な実験は3回のうち1回のみであった。30時間培養を続けた時点のプレートの画像を図11に示す。

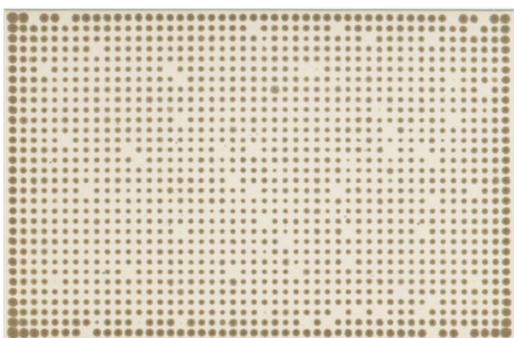
A) M9 Glucose



B) M9 Glycerol



C) M9 Succinate



D) M9 Xylose

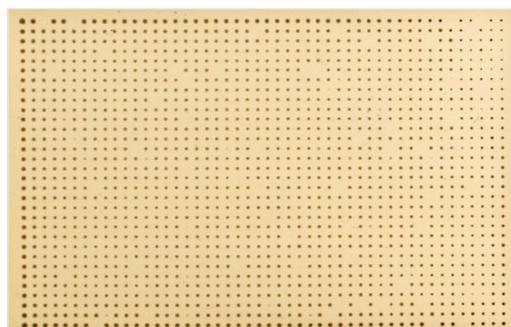


図11. 4種類の糖源によるM9寒天培地上での大腸菌コロニーの画像

A) M9 Glucose, B) M9 Glycerol, C) M9 Succinate, D) M9 Xyloseの各プレート上での大腸菌コロニーの生育を示す。最外周は全て野生型大腸菌を配置しており、一部コロニーが見えないところは、全遺伝子を網羅するのに必要な3枚のプレートを区別できるようにしたブランク。内側のブランクに関しては、その糖源の条件では生育できない欠失株がコロニーを形成でいなくなったためコロニーの生育が見られない欠失株。

M9 Xyloseを除く得られた3回の生育データの再現性を、それぞれの実験間の相関及び散布図で示した（図12）。その結果、全ての実験間で相関係数が0.8付近であることを確認し、再現性の確認が出来たと考えた。

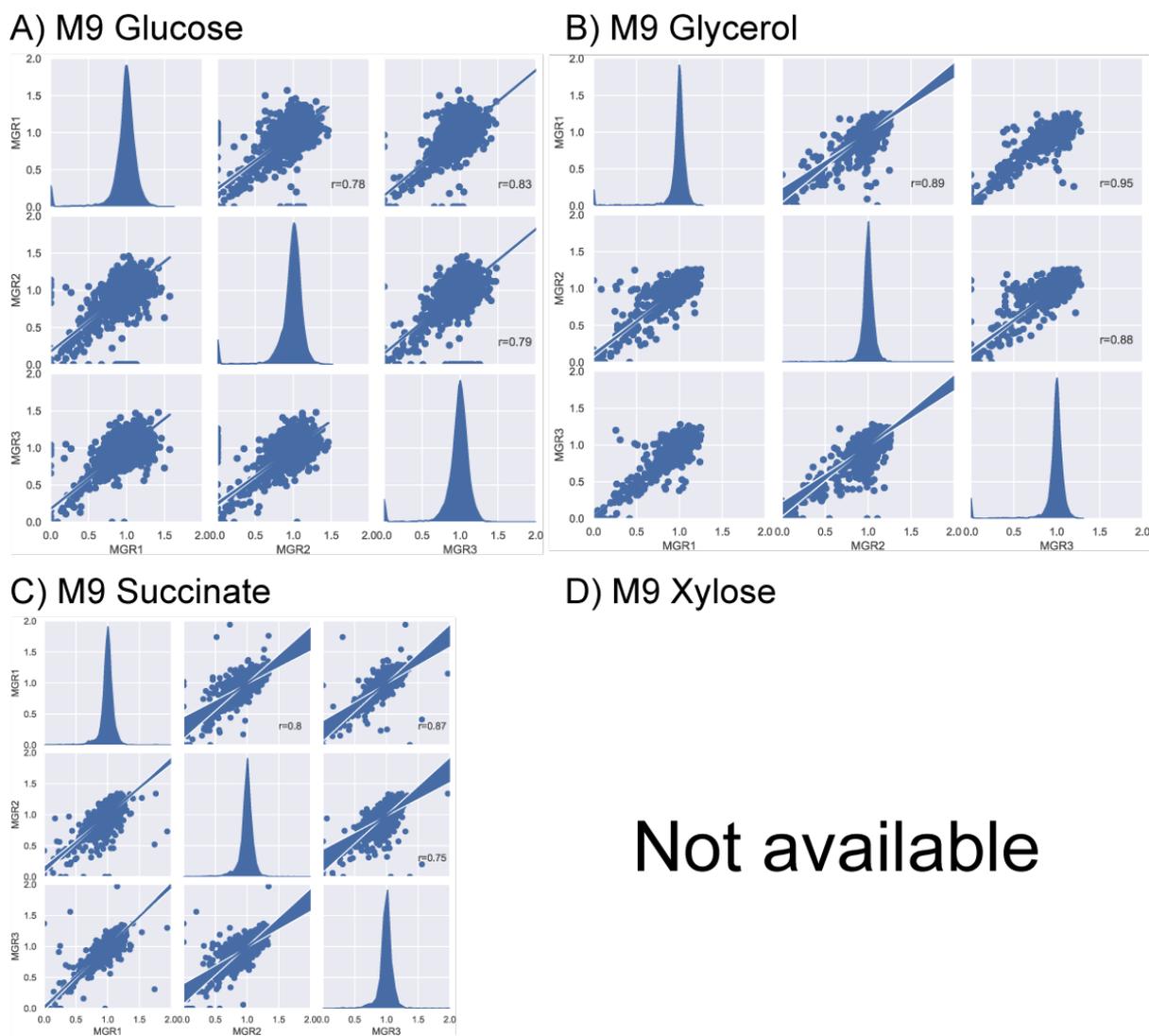


図12. 一遺伝子欠失株による4種類の糖源による生育の再現性

A) Glucose、B) Glycerol、C) Succinate、D) Xyloseの各糖源を加えたM9培地での生育をColony-liveを使用して3回観測したデータの相関を散布図で示す。D)のXyloseは、生育が悪くColony-liveでの定量は1回できたのみだったため、再現性の解析対象外とした。

本研究では、それぞれの実験で得られた生育度の中央値を計算し欠失によりMGRが0を示した遺伝子を致死を示す遺伝子として扱うこととした。欠失により致死を示す遺伝子の数を表5に示す。また、致死に至った遺伝子をGene ontology analysis

tool(GOAT) (Klopfenstein, D.V et al., 2018)を用いて機能分類を行い、各機能に分類される致死性を示す遺伝子の分布を調べた。なお、GOATは人やハエなどの解析を想定しておりそのままでは大腸菌には使用する事が出来なかったため、Gene ontology 公式サイト(<http://geneontology.org/>)から大腸菌のGOに関する情報ファイル(ecocyc.gafのv.14.1)をダウンロードし、GOATのプログラム内でダウンロードしたファイルを直接指定する事で解析を行った。機能別分類の結果、欠失した際に致死に至った代謝関連の遺伝子のうち、約6割がアミノ酸代謝を担う遺伝子、約2割が補酵素の生合成に関する遺伝子であった。

表5. 各糖源でのM9培地における致死遺伝子の分布状況

COG name	COG	M9 Glucose		M9 Glycerol		M9 Succinate		M9 Xylose	
			%		%		%		%
Transcription	288	3	0.01	8	0.03	13	0.05	3	0.01
Translation	170	0	0	4	0.01	1	0	0	0
Signal Transduction	193	0	0	2	0.01	8	0.03	1	0
Secondary Structure	95	0	0	1	0	0	0	0	0
Replication and repair	216	2	0.01	7	0.02	8	0.03	2	0.01
Post-translational modification protein turnover chaperone functions	133	2	0.01	4	0.01	5	0.02	1	0
Nucleotide metabolism and transport	84	16	0.06	14	0.05	18	0.06	16	0.06
Lipid metabolism	96	0	0	3	0.01	8	0.03	1	0
Intracellular trafficking and secretion	118	0	0	0	0	0	0	0	0
Inorganic ion transport and metabolism	253	8	0.03	14	0.05	12	0.04	6	0.02
General Functional Prediction only	446	2	0.01	15	0.05	15	0.05	2	0.01
Function Unknown	285	1	0	13	0.05	10	0.03	2	0.01
Energy production and conversion	282	3	0.01	8	0.03	25	0.09	12	0.04
Coenzyme metabolism	146	17	0.06	19	0.07	15	0.05	21	0.07
Cell wall/membrane/envelop biogenesis	219	1	0	9	0.03	5	0.02	1	0
Cell motility	152	0	0	7	0.02	8	0.03	0	0
Cell cycle control and mitosis	34	0	0	1	0	0	0	0	0
Carbohydrate metabolism and transport	358	1	0	17	0.06	14	0.05	4	0.01
Amino Acid metabolism and transport	391	52	0.18	64	0.22	74	0.26	56	0.19

致死に至った遺伝子を全ての糖源で致死を示した遺伝子、4つの糖源で培地特異的に致死を示した遺伝子の5グループに分け、Kyoto Encyclopedia of Genes and Genomes (KEGG) の機能であるKEGG Mapper v4.3(Kanehisa M., 2020)を用いた代謝経路へのマッピングを行い糖源による代謝経路上の致死遺伝子の違いを確認した。5つのグループのそれぞれの遺伝子の数を図13のベン図に、遺伝子名を表6に示す。

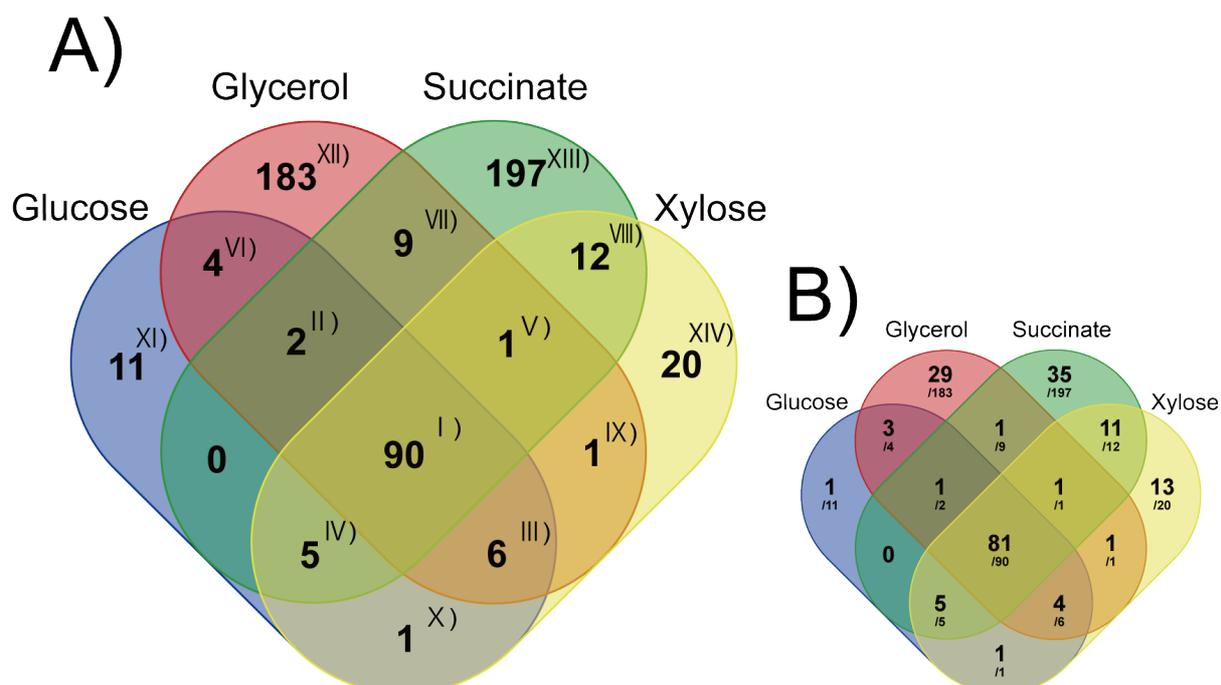


図13. 各糖源を添加したM9培地で致死性を示した遺伝子の数の分布をベン図で示した
A) 致死性を示した全遺伝子数、B) 致死性を示した遺伝子で、KEGGに登録されている遺伝子数。

表6. 各糖源を添加したM9培地で致死性を示した遺伝子

	遺伝子
I)	<i>argA argB argC argE argG argH aroA aroB aroC aroD aroE carA carB cysB cysC cysE cysG cysI cysJ cysN cysW exoX folB glmY glnA glyA guaA guaB hisA hisB hisC hisD hisG hisH hisI icd ilvA ilvC ilvD ilvE iscS ldcA leuA leuB leuC leuD lpd lysA metA metB metC metE metF metL metR nadC oxyR pdxH pheA proA proB proC purA purC purD purE purH purK purL purM pyrE pyrF serA serB serC thiC thiD thiE thiF thil thrA thrB thrC tpiA trpA trpB trpC tyrA ycaL ygfB</i>
II)	<i>purF leuL</i>
III)	<i>leuL lipA hisF thiG thiH thiS</i>
IV)	<i>cysD cysH nadB pyrB pyrC</i>
V)	<i>pgpA</i>
VI)	<i>fepB fepC fepG nadA</i>
VII)	<i>ccmD hdeB mepA phoU proQ yaiA ycjV yfiL yjeI</i>
VIII)	<i>atpE atpF atpG atpH cyaA cysK dksA gltA ubiE ubiG ubiH xylA</i>
IX)	<i>ppc</i>
X)	<i>pyrD</i>
XI)	<i>dapF dnaK fepD fes hflD holC pfs prc yccE ycdY ydaT</i>
XII)	<i>agal alsA alsE ampE araF aroL astA atoB bcsF betT C0664 cdaR clcB cof corA cpdA cri crr cspD cspG cynT damX dedD dkgB dusC emrB entB fecD fepE fhuC fhuF fieF fimI fliA fliR frc friB fsaA galT glcE glnB glnE glxR gshB gspC hda hdeA hokC hrpA hslR htrC hupB hybA hyfH I004 kdgT ligB livH livJ lplA menF mfd moaB motA mtlD nagZ nanE ndk nika nlpA nrfA nrfB ompC ompN pepT pfkA pheP pin pncA pnp pnuC potA pqiA rbbA relE rhsB rna rpsF rrmA ryeA sbmC sgbH sgrS slyB sodA sseB stfE sufB sufE topB torD TU-3697 TU-5237 ubiX ugpB ulaD umuC usg yahH yaiU yajQ ybaQ ybcJ ybcV ybdK ybeQ ybfH ybgL ybgQ ybiU yccA ycdB ycdO ycgL ycgN ydcN ydeM ydfG ydfR ydgD ydhF ydiN ydjl ydjN yeaJ yeaT yebF yecM yedD yeeP yegX yeiM yfcM ygaY ygbA ygeA ygfK yggU ygiL ygjR yhfA yhfS yhhN yiaN yicH yicM yidX yieP yifE yihN yiiX yjaG yjbL yjbN yjbQ yjdL yjeO yjfC yjgB ykgB ymdF ymfP ynal yohJ ypdI ypfH yphB yqeG yqgE yqhC ytfG zur zwf</i>
XIII)	<i>aceA acrR agaR ahpC alpA ampC apaG argD argI aslA asmA atpA atpB barA beta bioA bisC chbA chbB cheB cmk cobC copA crp cybB cysP dcd dgoA dgoT dhaL dinG dppB dusB etp eutG fabF fbp fecl fhlA fimF flgF flhA folM folX galP glpQ gnsB gph greB grxB hflX hisP hscB intS iscU IstR-2 lldP malG mdtA menC modB mog mpaA mtlR nanK narZ nei nfo nrfE nuoL paaD pabA pepP pgpB phnF phoE php pncB pppD priC puuB rcsF rspB ryfA sbcD sdhA sdhB sdhC sgcA slit sroH sspB ssuE trkA ttk TU-1150 TU-3687 tufA ubiF uidR yacF yafW yagB yahA yahN ybdD ybeL ybgJ ybgK ybjL ybjS yccF yccS ycdC ycdH yceG ycfP ychJ yciA ycil ydaO ydaV ydbD ydcM ydcU yddB yddG ydfA ydfI ydfZ ydgH ydhC ydhY ydiD ydiV ydjG ydjJ ydjO yeaO yebA yebB yecG yedV yeel yeeY yegl yehQ yehY yeiP yfaX yfcL yfdN yfdY yfeH yfeK yfeY yfgH yfiA yfjT ygaC ygcF yggX yghA yghE yhbJ yhbQ yhbT yhdW yhgA yhiN yihG yjeM yjeT yjgl yjhQ yjiO yjiW ykiA yncC yneF yneJ ynfA yniC yoaA yoaH yobD yoeB yohN ypjB ypjM yqaB yqcE yraM yraR yrhA ytfF ytfL</i>
XIV)	<i>aceE aceF cysQ glnP hns ihfA lipB pal panC panD pdxA pdxB pdxJ rfaE rpe speE xylB xylR ycbK ygfZ</i>

図13から、90の遺伝子が全条件で共通に致死性を示す事が明らかになった。また、そのうちKEGGにMappingされた酵素遺伝子は81個、その中でアミノ酸代謝関連遺伝子は51個であった。次に、それぞれ培地特異的に致死性を示した遺伝子をKEGG PATHWAYの経路情報を参照しMappingした結果、M9 GlycerolとM9 Succinateは殆どがマッピングされなかった。しかし、M9 SuccinateではSuccinateからFumarateへの経路を担う*sdhA*、*sdhB*、*sdhC*の欠失株が致死性を示していた。M9 Xyloseに関して大半はM9 Glucoseと共通の遺伝子で致死を示していたが、Xyloseからペントースリン酸経路までを担う3つの遺伝子のうち*xylB*、*rpe*の2つが独立で致死性を示し、*xylA*はM9 SuccinateとM9 Xyloseのみで致死性を示していた。

次に、定量実験で致死に至った遺伝子とClusters of Orthologous Groups (COGs) (Galperin et al., 2015)の情報を照らし合わせ、どの機能をもつ酵素を担う遺伝子が多いか確認を行った。それぞれの実験条件の結果を図14に示す。X軸の数値は、それぞれのCOGsグループ中の遺伝子の総数からみた比率を表している。図14A)~D)から、全ての条件でAmino acid transport and metabolism、Nucleotide transport and metabolism、Coenzyme transport and metabolismの3グループがそれぞれ全体の2割を占めており、代謝と取り込みに偏りを示す事が分かった。

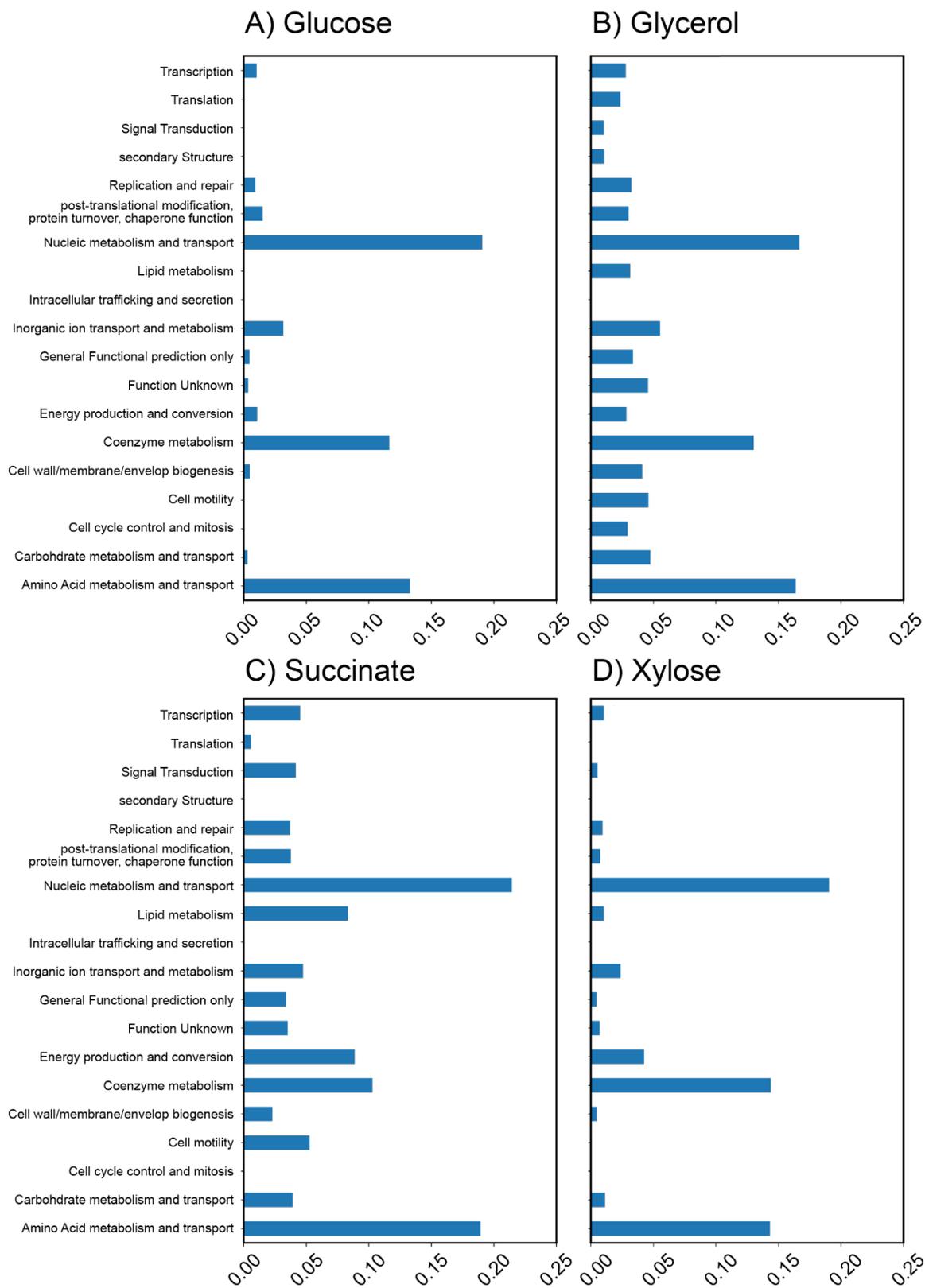


図 14. A)~D)各糖源を添加した M9 培地で致死性を示した遺伝子の COGs id を用いた機能別分類

2.3.2 シミュレーションと実測との間に食違いのある遺伝子群

定量実験で得られたMGRの値と、シミュレーションで計算した予測値からFNを示す遺伝子の同定を行った。シミュレーション実験での生育スコアの閾値は定量実験と同様にMGRが0かそうでないかとし致死と予測された遺伝子の数を算出し(表7)、定量実験とシミュレーション実験の比較を行った。

表7. モデルに組み込まれている遺伝子群の各糖源で致死性を示した遺伝子数

	M9 Glucose	M9 Glycerol	M9 Succinate	M9 Xylose
iJO1366での致死数	208	219	215	209

iJO1366では、1366の大腸菌遺伝子がモデルに組み込まれており、その中での致死性を示した遺伝子数。

比較した結果FNを示した遺伝子名を表8に、またFNを示した遺伝子の各糖源間での重複をベン図として図13に、表として表9に示す。図13よりM9 GlucoseでFNとなった遺伝子は全て、他のいずれかの培地条件と共通であることが分かった。

表8. それぞれの炭素源でFalse negativeを示した遺伝子リスト

培地	FNを示した遺伝子群
M9 Glucose	<i>aldA, argD, aspC, bioA, bioB, bioC, bioD, bioF, bioH, cyaY, cysQ, fabH, gltA, kdsC, luxS, moaA, moaC, moaD, moaE, mobA, moeA, moeB, mog, pabA, pabB, pabC, panB, panC, panD, pdxA, pdxB, pdxJ, trpD, trpE, ubiC, ubiX, yrbG, zupT</i>
M9 Glycerol	<i>thrC, mog, pdxA, leuC, leuB, leuA, panD, panC, panB, thil, purK, purE, gltA, bioA, bioB, bioF, bioC, bioD, moaC, moaD, moaE, moeB, moeA, serC, aspC, pyrC, fabH, pabC, icd, trpA, trpB, trpD, trpE, pyrF, aldA, aroD, pabB, hisG, hisD, hisC, hisH, thiD, ubiX, pdxB, aroC, purM, guaA, iscS, purL, pdxJ, nadB, luxS, cysC, cysN, cysD, cysH, cysl, cysJ, argA, lysA, zupT, yrbG, kdsC, argD, pabA, cysG, bioH, cysE, atpC, atpD, atpG, atpA, atpH, atpF, atpE, atpB, ilvE, ilvD, ilvC, cyaY, dapF, mobA, tpiA, metF, argC, argB, argH, thiH, thiG, thiF, thiC, purD, purH, metA, ubiC, purA, cysQ, pyrB</i>
M9 Succinate	<i>thrB, thrC, mog, pdxA, leuD, leuC, leuB, nadC, panD, panC, thil, purK, purE, gltA, sdhC, sdhD, sdhA, sdhB, nadA, bioB, bioF, bioC, bioD, moaC, moaD, moaE, moeB, moeA, aspC, pyrC, fabH, pabC, icd, trpA, trpB, trpC, trpD, trpE, pyrF, aldA, aroD, pabB, hisG, hisC, hisB, hisF, thiD, ubiX, pdxB, aroC, purM, guaA, iscS, purL, pdxJ, nadB, pheA, luxS, cysC, cysN, cysD, cysH, cysl, lysA, zupT, folB, yrbG, kdsC, pabA, cysG, bioH, cysE, pyrE, ilvE, ilvD, ilvC, cyaY, dapF, mobA, tpiA, argE, argC, argB, thiH, thiG, thiC, purD, purH, metA, ubiC, cysQ, thiS</i>
M9 Xylose	<i>mog, panB, nadA, bioA, bioB, bioF, bioC, bioD, moaA, moaC, moaD, moaE, moeB, moeA, aspC, fabH, pabC, trpD, trpE, aldA, pabB, ubiX, purF, luxS, zupT, yrbG, kdsC, argD, pabA, bioH, cyaY, dapF, mobA, ubiC</i>

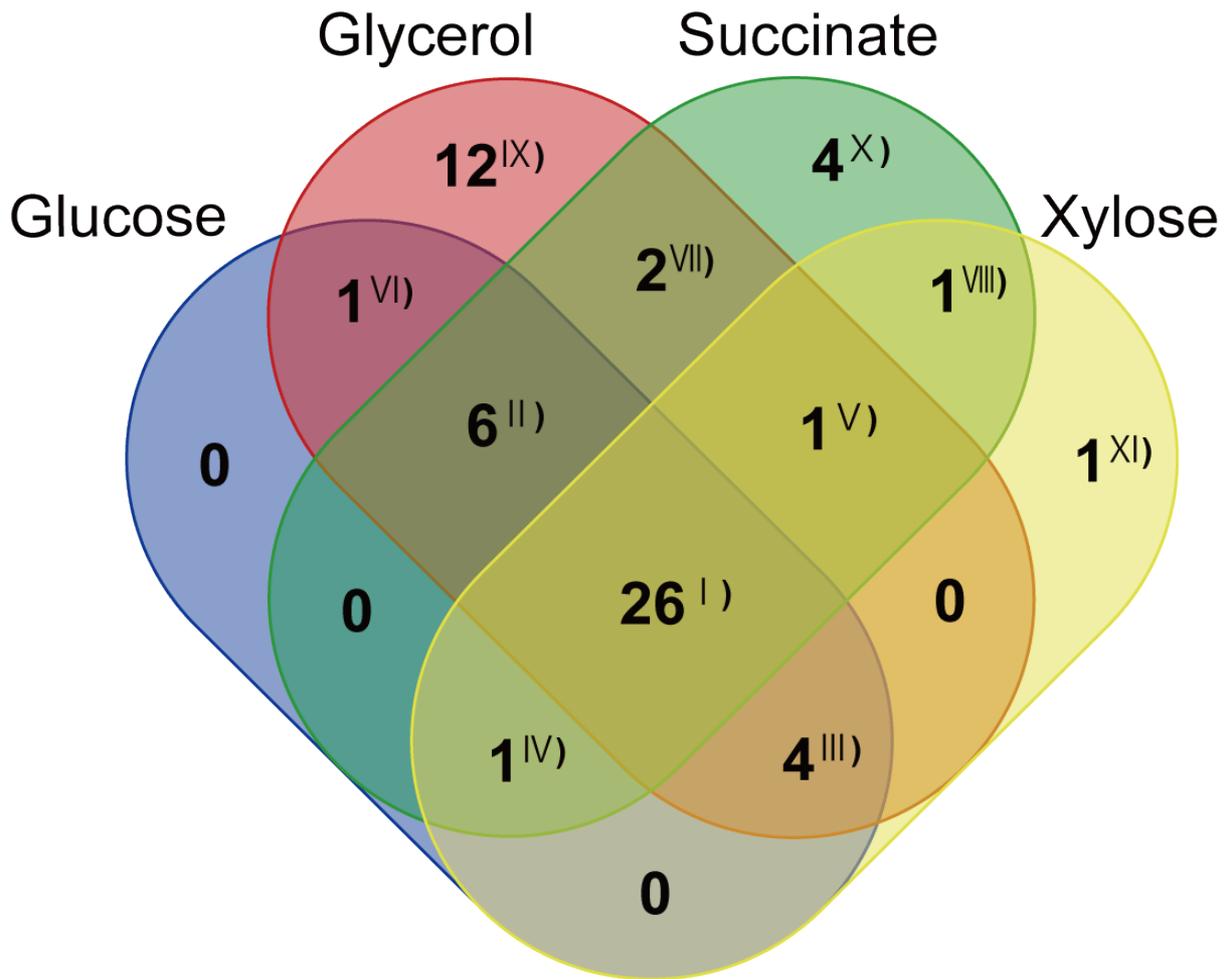


図15. 各糖源を変えた最少培地においてシミュレーションと実測で食違いを示す遺伝子数分布

表9. それぞれの糖の組合せで共通に観察されたFalse negativeを示した遺伝子

	total	分類された遺伝子
I)	26	<i>fabH, bioF, zupT, moaA, trpD, mobA, moaC, trpE, bioD, panB, moeA, pabB, ubiC, bioC, luxS, moaD, moeB, aldA, cyaY, moaE, bioB, pabC, kdsC, yrbG, aspC, bioH</i>
II)	6	<i>panD, pdxA, cysQ, pdxB, pdxJ, panC</i>
III)	4	<i>argD, pabA, bioA, mog</i>
IV)	1	<i>ubiX</i>
V)	1	<i>dapF</i>
VI)	1	<i>gltA</i>
VII)	2	<i>thiH, nadB</i>
VIII)	1	<i>nadA</i>
IX)	12	<i>cysH, atpH, atpF, atpC, pyrB, atpB, cysD, atpG, pyrC, atpA, atpE, atpD</i>
X)	4	<i>hisF, thiS, thiG, sdhD</i>
XI)	1	<i>purF</i>

2.3.3 最少培地を用いた二重欠失株作製法の確立

最少培地を用いた二重欠失株作製による遺伝的相互作用解析を進める為、M9最少培地による二重欠失株作製の効率を上げる手法の確立を目的として、手法の改善に取り組んだ。

接合効率に影響が考えられるステップは図9の通り1つ目は、Fプラスミドの*oriT*にニックが入り、染色体複製を伴い、受容菌へ移動するDNA移動効率である。2つ目は、Type IV secretion systemを祖先型に持つ接合機構が膜上に構築され、受容菌を捕捉し、細胞質間の連絡チャンネルが形成される段階である。3つ目は、挿入された染色体断片が受容菌染色体に相同組換えによって挿入される組換え効率である。当初は、微量代謝産物の環境中からの持ち込みが、その後の細胞の生育に影響を及ぼす事が明らかにされているため、全てのステップをM9最少培地で行うことを検討したが、図16に示す様に、コロニーのムコイド化やコロニー形状の不均一性など多くの問題が存在した。

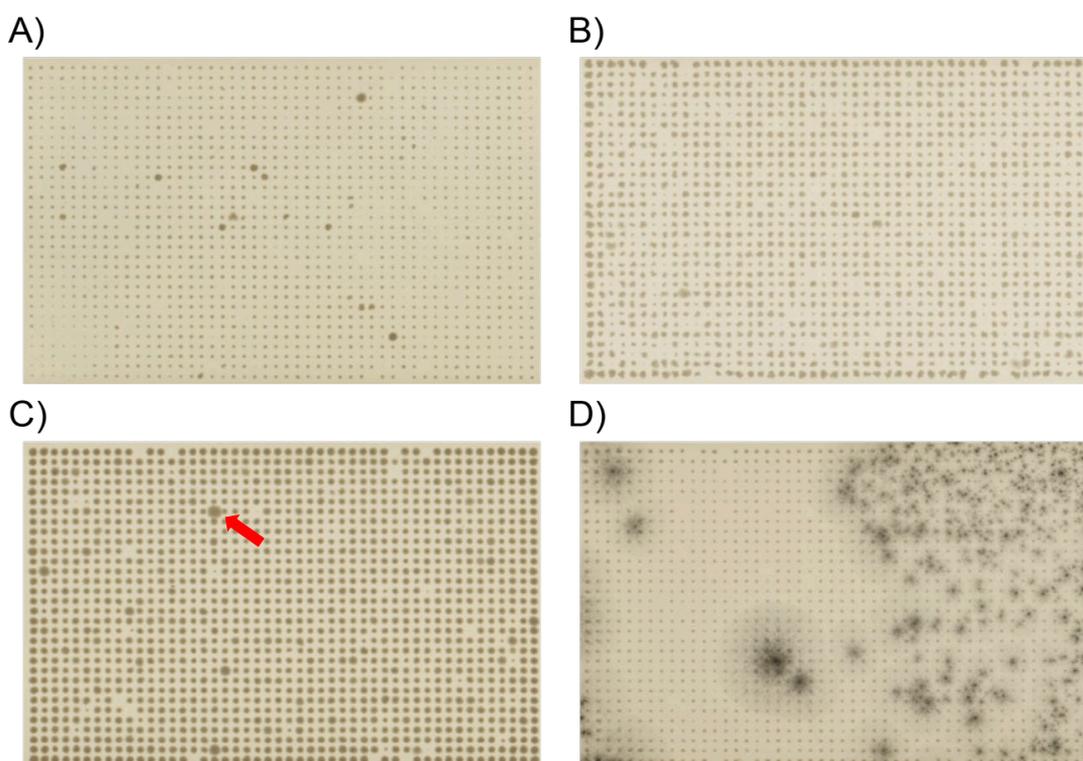


図16. 全ての工程をM9培地で行なった場合の生じた問題

A) きわめて一部しか二重欠失株が作製されておらず、二重薬剤による選択でほとんどコロニーが形成できない。B) 二重欠失株作製の効率が悪く、非常に少ない細胞のみが培地上で増殖した結果、コロニーを円状に近似する事が出来ず歪な形状を示す。C) 一部の細胞(赤矢印の箇所)はムコイド状の形態を取り、定量が難しい。D) 長時間の培養によるカビ等の混入。

以上の問題を解決する為、表10に示す様に、検討すべきステップを前培養、接合、スクリーニング、そして定量化の為のコロニー形成モニターに分け、それぞれのステップにおいて、培地条件及び時間、必要性、無菌環境維持手法の検討を行なった。

表10. 二重欠失株作製方法改善に向けて検討すべき項目と内容
スクリーニング

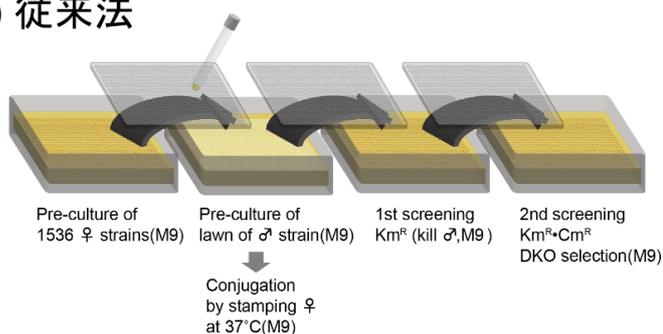
ステップ 対象	前培養	接合	スクリーニング			生育モニター
	培地	培地・時間	1次 培地・時間	2次 培地・時間	3次 必要性	無菌環境構築
内容	LB M9	LB X 時間 M9	LB X 時間 M9	LB X 時間 M9	時間	

従来は全ての工程をLB培地で行なっていたが、M9最少培地による解析を行うために、全工程をM9最少培地に変更したところ、図16の様な問題が生じた為、接合伝達による二重欠失株の構築まではLB培地で行い、薬剤による選択の段階で最少培地を用いる方向で検討を進めた。LB培地で接合を行う事は、これまでに確立してきている手法であり、確実に二重欠失株を取得できる。一方、選択の段階で組換えが起こり、二重欠失株が作製されるが、この段階の効率を上げることと、観測段階においてLB培地の栄養の持ち込みによる影響を低減させる必要がある。以上の点を考慮し、薬剤スクリーニング回数を増やす事を検討した。組換えにかかる事の出来る時間を伸ばす事で、より多くのDKO株を得る事が出来ると考えられる。併せて、薬剤投与のタイミングを変え薬剤によるストレスを抑えることで組み換え効率の低下を防ぐ事が出来ると考え検討を行なった。

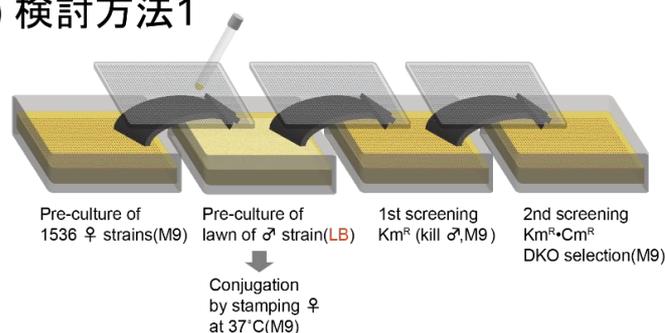
最初に、供与菌となるHfr株をM9 Glucose液体培地ではなく栄養培地であるLB液体培地で前培養したものを接合に使用し、複数回DKO株構築及び生育観測実験を行った。次にHfr株及び受容菌のKeio collectionを、栄養培地上で行う事でDNA移動効率及び接合機構確立効率が低下しないことを期待し、前培養から接合までLB培地を使用し接合させて複数回生育観測を行った。次に、接合後の薬剤による選択段階の検討を進めた。これまでの方法では、接合後に供与菌を殺し、次の段階で二重欠失の

選択を行いながらモニターを行う2回の薬剤を用いた選択を取っていた。これを2回から3回に増やし薬剤投与の組み合わせも併せて変更する事で薬剤ストレスの低減および生育時間を延ばすことによる二重欠失株の取得率増加を期待し検討を行った。この際、薬剤投与のタイミングを2nd Screening、3rd ScreeningともにKm・Cmを用いた場合と2ndにCm、3rdにKm・Cmを用いた場合の2パターンに分けて検討を行った。検討を行なった実験方法を模式的に図17へ示す。

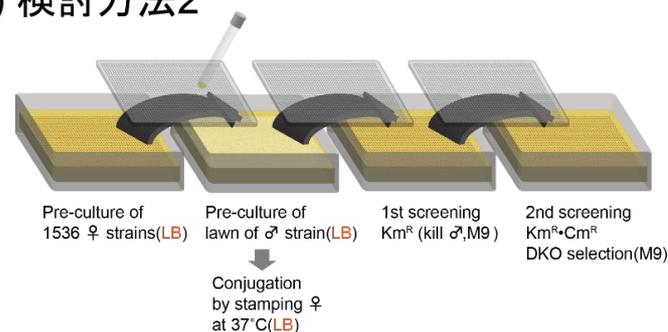
A) 従来法



B) 検討方法1



C) 検討方法2



D) 検討方法3

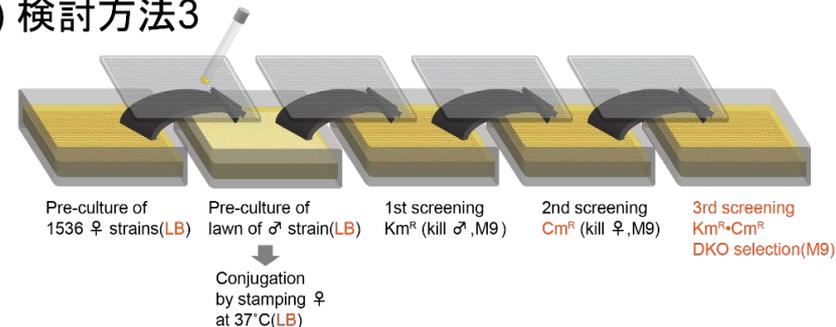


図17. 検討方法の模式図

従来の方法は、全てLB培地で行い、接合後の選択は2回で、2回目にスキャナーでのコロニー画像の取得を行ってきた。これに対して、B) 前培養をLBで行い、以降は全てM9培地で行い、2回の選択、C) 前培養と接合をLB培地で、選択をM9培地で2回、D) 前培養と接合をLB培地で、選択を1回目で供与菌を排除、2回目で受容菌を排除、3回目で二重変異株の選択を行い画像取得を行う方法を検討した。

図17に示すように、検討を行なった条件は、前培養でLBかM9培地を、接合を行う

際の培地条件をLBとM9培地で、接合後の選択を行う段階での選択方法、の3条件を検討した。

検討を行なった結果を表11及び図18に示す。

表11. 二重欠失株作製の条件検討の結果

培地条件	前培養	接合	選択培地	選択回数	実験数	画像取得数	画像取得失敗	定量化成功	定量化失敗	標準化失敗	欠損値なし	標準化成功数	定量化成功率
LB	LB	LB	LB	2	513	501	12	469	32	66	403	403	78.6
	LB	LB	LB	3	163	115	48	60	55	15	45	45	27.6
M9 Glucose	M9	M9	M9	2	206	134	72	85	49	72	13	13	6.3
	LB	M9	M9	2	222	174	48	146	28	138	8	8	3.6
	LB	LB	M9	2	507	484	23	406	78	272	134	134	26.4
	LB	LB	M9	3	364	316	48	279	37	15	264	264	72.5
M9 Glycerol	M9	M9	M9	2	132	58	74	43	15	43	0	0	0.0
	LB	M9	M9	2	48	44	4	33	11	33	0	0	0.0
	LB	LB	M9	3	236	216	20	159	57	24	135	135	57.2

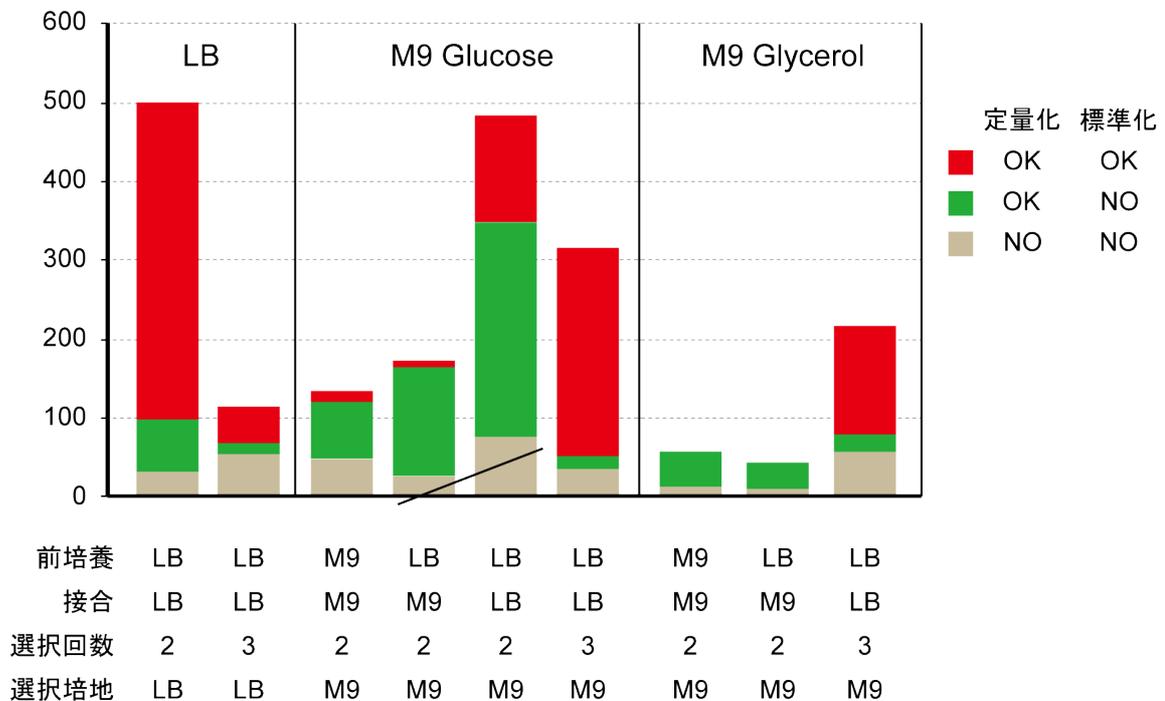


図18. それぞれの検討方法の定量化および標準化成功数

様々な条件下でのLB寒天培地とM9 GlucoseおよびM9 Glycerol寒天培地におけるDKO株観測実験での定量化および標準化処理の結果を示す。LBでは薬剤を用いた選択回数をふやすと定量化の時点で失敗する数が増えた。M9 GlucoseおよびM9 Glycerol寒天培地では、接合までLB寒天培地上で行い、薬剤による選択回数を3回へ増やす事で標準化成功まで到達する数が大きく増加した。

これまでの様にLB培地だけでの遺伝的相互作用解析では、選択方法を2回から3回に増やすと、定量の成功率が78.6%から27.6%へと大きく低下してしまった。一方、最少培地で網羅的遺伝的相互作用解析を行なった場合、選択方法を2回で行うと、M9 Glucoseで26.4%と成功率が非常に低い。M9 Glycerolに至っては0%と全く定量ができなかった。そこで接合後受容菌側に移動した供与菌染色体DNAと受容菌染色体DNAの間に相同組換えが起こる(図9)効率の違いを想定し、選択方法の検討を行なった。1回目の選択で供与菌を排除し2回目で組換え体を選択するという方法から、2回目に受容菌のみを排除する選択を追加し、その後3回目に組換え体を選択す

るクロラムフェニコールとカナマイシンの二重選択を行うステップを導入した方法へと変更した。その結果、M9最少培地を利用した二重欠失株作製での成功率が、M9 Glucoseで72.5%、M9 Glycerolでは57.2%へと改善が見られた。(表11)

実験の過程で、図17D)に示すプロトコルの再現性を見るために複数回確認を行なった結果を図19に示す。*eda*, *cyaA*, *yciG*, *gltA*, *kdsC*の5遺伝子について2回の実験を行い、それらの再現性を見たところ、もっとも高い再現性を示したのは $r=0.79$ の*eda*であった。もっとも低いものでも、 $r=0.587$ であり再現性を確認する事が出来た。

以上の結果をもって、前培養と接合をLB培地で、接合後の選択を1回目はカナマイシンを加えたM9培地で、2回目はクロラムフェニコールを加えたM9培地で、そして3回目にはカナマイシン及びクロラムフェニコールを加えた培地での選択を行い、解析を進めることとした。

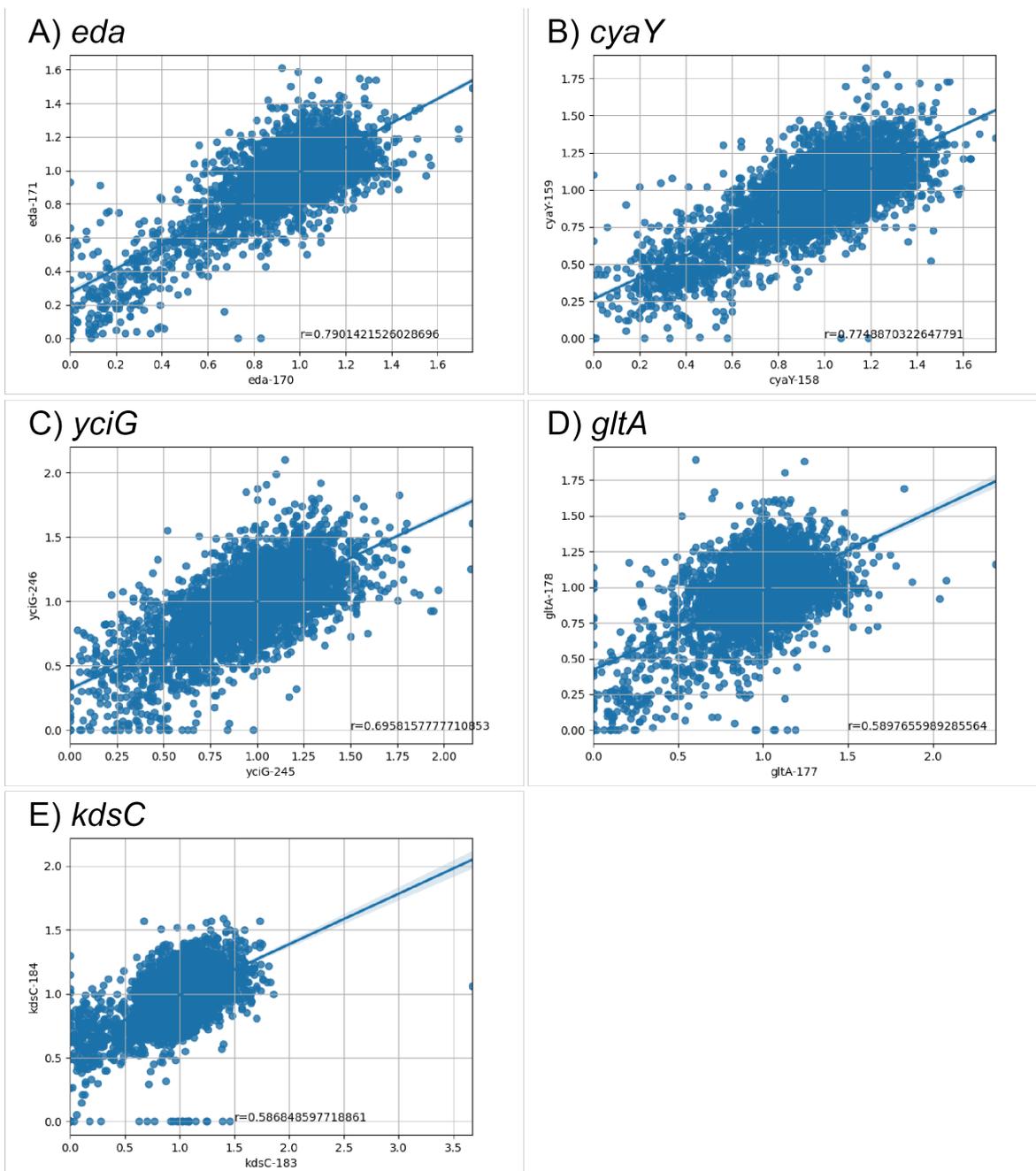


図19. 改善方法による網羅的遺伝的相互作用解析の再現性

図上に示した遺伝子を供与菌として、その欠失を受容菌に移動させ二重欠失株を作製し、改善方法に従い、選択を3回行った場合の定量値の再現性を散布図で示す。

2.4 考察

2.4.1 一遺伝子欠失実験及びシミュレーションとの比較

4糖源(Glucose、Glycerol、Succinate、Xylose)を炭素源としたM9最少培地を用いて、Keio collectionの一遺伝子欠失株を用いた全ての遺伝子欠失株による生育の実測値と、iJO1366モデルを用いたFBAによるシミュレーション実験結果とを比較を行い、違いを示す遺伝子欠失株の同定を行った。

30時間培養後のコロニーの生育を寒天培地の写真イメージを図11に示したが、M9 Xyloseを糖源とした培地での大腸菌細胞の生育が、他と比較して生育が非常に悪い。そこで野生型大腸菌のそれぞれの糖源での生育の実測値とシミュレーションによって計算されたMGRの値を表12に示す。

表12. 糖源ごとの生育の実測値とシミュレーションで予測された生育速度

	MGR	SPG	Simulation
Glucose	0.03	110.87	0.98
Glycerol	0.05	69.21	0.21
Succinate	0.05	109.70	0.81
Xylose	0.01	12.42	0.49

iJO1366モデルによるシミュレーション結果では、M9 Xyloseで0.49と予測されたが、実測での生育速度（MGR）は、Glucoseは0.03、Glycerolでは0.05、Succinateでは0.05に対して、Xyloseでは0.01と生育速度の低下が著しい。Colony-liveでは最大到達濁度(SPG)の計算も行なっているが、その値は、Glucoseで110.87、Glycerolで69.21、Succinateで109.70に対して、Xyloseでは12.42とGlucoseと比較して約1/10にまで下がる。Glucose、Glycerol、Succinateを糖源とした場合には、良好な生育を示し定量化も問題はなかったが、Xyloseでの定量は3回の実験で1回しか定量ができず、再現性の確認を行う事ができなかった。Xyloseで生育の定量ができなかったのは、他の糖源と比べて生育が非常に悪かった為であるが、画像のアンシャープ化に使用する手法とパラメーターの調整、画像の二値化処理の閾値の調整により画像解析でコロニーのエッジを検出する段階を調節する事でコロニーを特定し定量化まで進むことが可能であろうと考えられる。

次に、一遺伝子欠失株ライブラリーを用いた生育実験の結果、致死性を示した遺伝

子群をGOによる機能分類ツールのGOATにより解析を行った結果(表5)、ほとんどが代謝に関連していることがわかった。そして代謝関連に分類された遺伝子の約6割がアミノ酸代謝を担う遺伝子、同様の生合成に関連する遺伝子であったため、遺伝子欠失による代謝経路の遮断の影響を大きく受けていると考えられる。

致死性を示した遺伝子群をKEGG Pathwayへマッピングを行なった結果(図20)、全ての糖源で共通であった90の遺伝子(図13の1))はアミノ酸、核酸、硫黄代謝関連に分類される遺伝子であった。また培地特異的に致死性を示した遺伝子を同様にKEGG pathwayにマッピングした結果、GlycerolとSuccinateは培地特異的に致死性を示す遺伝子の数が多い。Succinateに関してはSuccinateからFumarateへの経路を担う *sdhA*、*sdhB*、*sdhC*(図21)が致死を示していた。

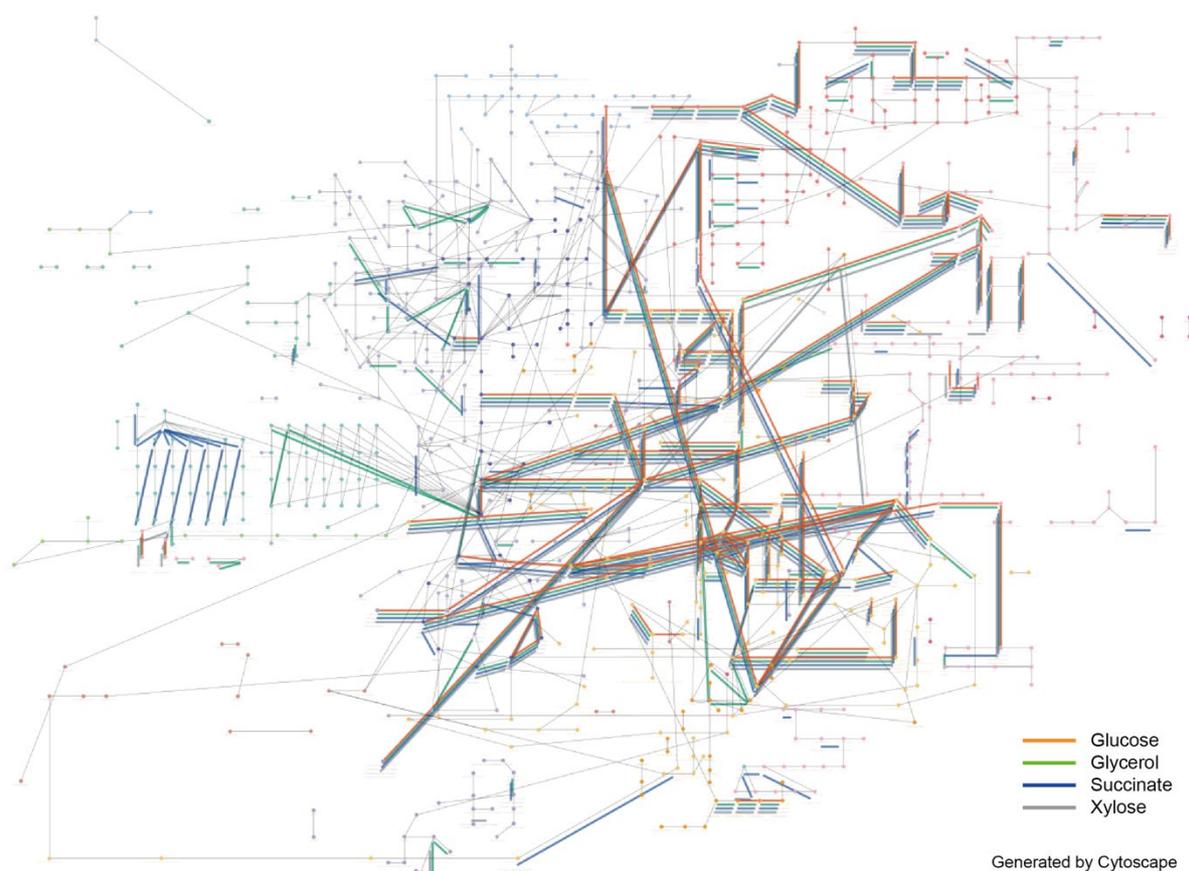


図20. 糖源の違いによるSKO株致死性の違い

オレンジの線はM9 Glucose、緑の線はM9 Glycerol、青の線はM9 Succinate、灰色の線はM9 XyloseでSKO実験において欠失により致死性を示した遺伝子が担っている経路を表す

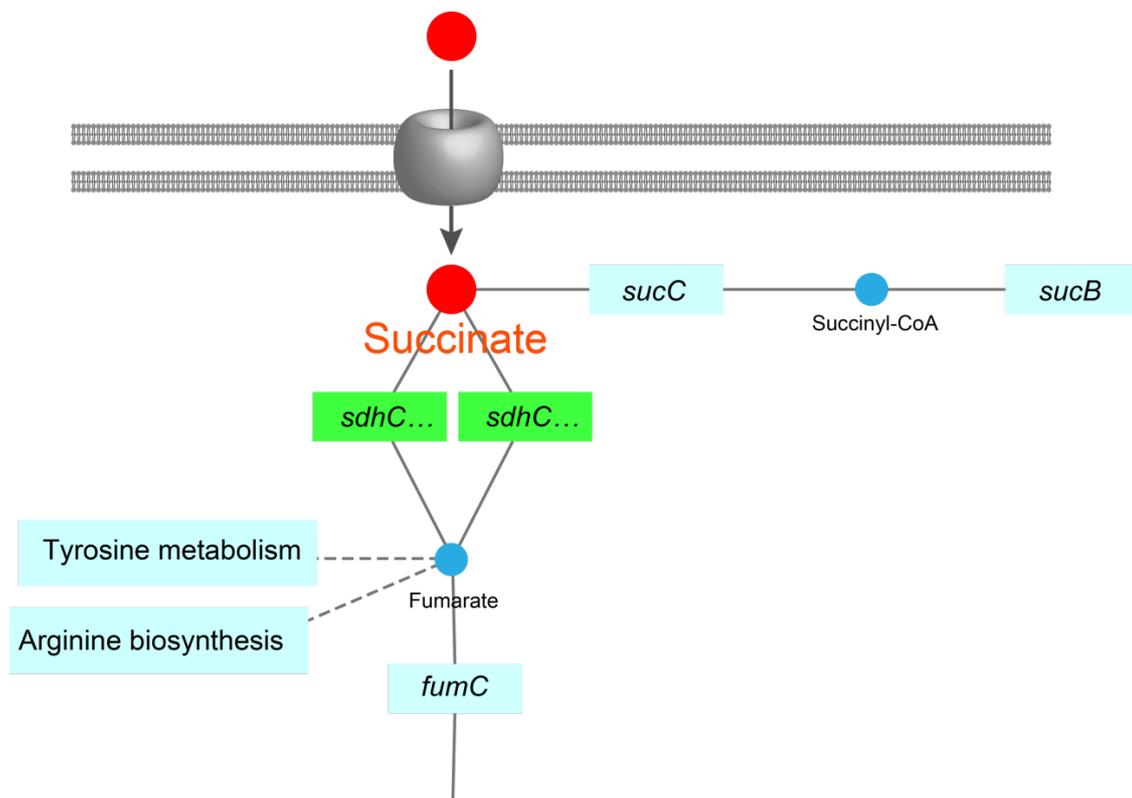


図21. Succinate 取り込み直後のステップでの致死性

また表6の13)より *atpA*、*atpB*の2つの遺伝子が独立で、その他の*atp*遺伝子群も表6の8)より Xyloseと共通であったが致死性を示しており、これはSuccinateに関する過去の文献で同様の報告がされている事がわかった(Boogerd et al., 1998)。培地特異的に致死性を示す遺伝子のうちGlycerolでは73、Succinateでは97の遺伝子が機能未知もしくは機能の予測のみがアノテーションされているものであった。よって、機能未知がGlycerolとSuccinateで致死だった遺伝子グループの大多数を占めている事から現状では既知の機能情報を用いたエンリッチメント解析等の特徴づけは難しいと考えられる。Xyloseに関しては、Xyloseからペントースリン酸経路へ入る手前までの経路を担う遺伝子が3個中2個独立で致死性を示し、1個がSuccinateと共通で致死性を示した(図22)。

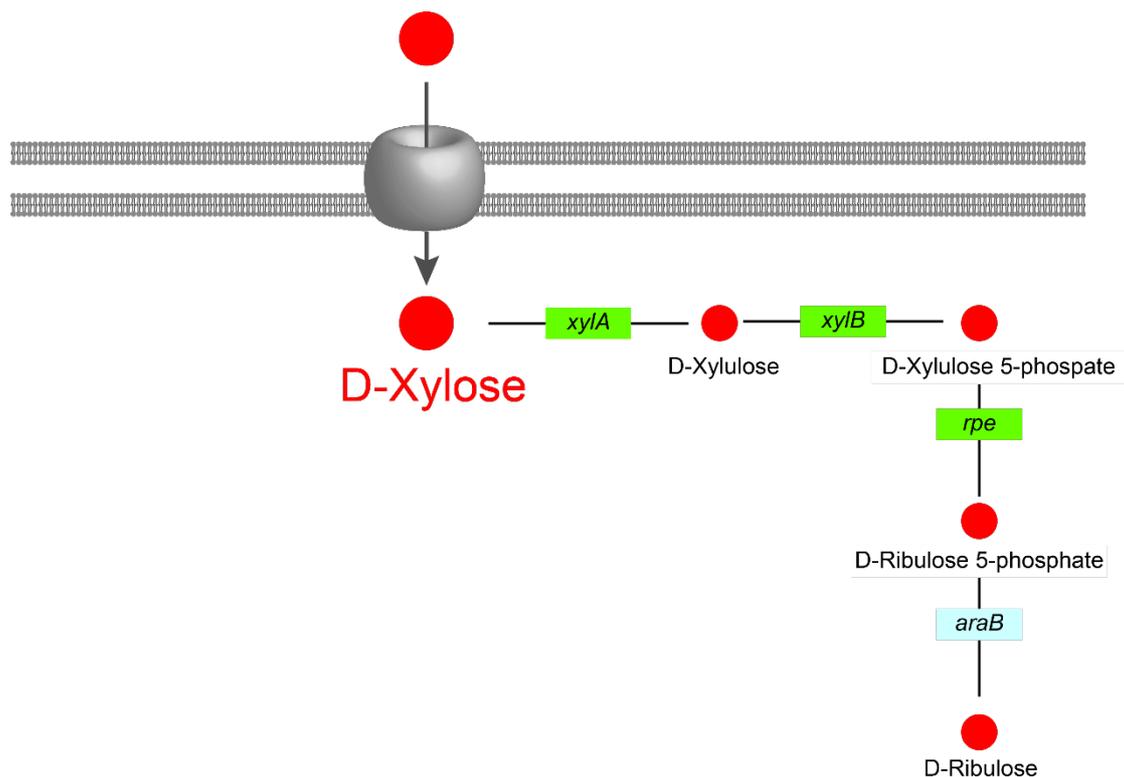


図22. Xylose取り込み直後の致死性を示すステップ

KEGG上では、Xyloseからペントースリン酸経路の手前までの経路には分岐がなく代替経路が存在しておらず(図22)、この経路が遺伝子欠失によって遮断された結果、Xyloseが取り込まれた後代謝が進まず致死に至った結果が表れたと考えられる。以上より、LB培地ではなく糖を加えた最少培地を用いた事により糖を取り込んだ後に働く直近の経路が培地特異的に致死を示しており、代謝経路における必須性を反映した結果が得られていると考えられる。

次に、今回の最少培地を用いたSKO実験と同じ条件に設定したシミュレーション実験を行い、FNを示す遺伝子を抽出した。その結果、M9 GlucoseでFNとなった遺伝子は全て、他のいずれかの培地条件と共通であった。よって、M9 Glucose上でFalse negativeを示した遺伝子に対して実験と解析を行う事で他の糖源でFNを示した遺伝子の検証も出来ると考え、M9 GlucoseでFNを示した遺伝子とKeio collectionとを組み合わせた網羅的な二重遺伝子欠実験による、本研究の目的であるFNを示す遺伝子を代替する遺伝子の探索を行うこととした。

2.4.2 最少培地を用いたDKO株作製プロトコルの改変

当初、当研究室で確立されたLB培地を用いたDKO株生育観測実験法(図10)を用いて、FNを示した遺伝子をHfr化しKeio collectionと組み合わせM9培地を用いたDKO株生育観測実験を行おうとしたが、ほとんどの株が生育せず、頻繁にカビが発生してしまう結果となった。カビに関しては、生育時間の長さから測定時間を長くする必要があり、さらにスキャナーへのプレートの設置は蓋を外しての設置であるため、浮遊胞子が混ざる危険性は高い。さらに恒温培養装置内のファンによる空気の循環も胞子の混入を促進すると思われる。そこで対策として図23のようにインキュベータ内部に取り外し可能なUVランプを設置し、観測を行う前にUV照射による滅菌工程を導入した。さらに、周辺環境も浮遊胞子等をできる限り除くことを目的に、UVランプによる滅菌をタイマーで制御することで定期的な実験環境の滅菌作業を行った。

A) 恒温培養装置内



B) 培養装置周辺



図23. UVランプによる恒温培養装置内及び設置環境の滅菌

小型の45cm長のUVランプを、庫内にはマジックテープで、必要時に設置しUV照射滅菌を行なった。実験室内には木ネジで常設固定しタイマーを使用して定期的に滅菌を行なった。

UV照射による滅菌を行うことにより、カビの発生は抑える事ができ、30時間を超える生育画像取得にもカビの生育が見られなくなった。

生育が出来ない、もしくはコロニーの形状がいびつになる原因としては、接合による欠失部位の挿入の効率が最少培地での条件では悪くなり、染色体DNAの接合移動後の相同組換えによりDKO株になる大腸菌の数が少なかったため円に近似した形へ

とコロニーを形成出来なかったことが考えられる。そのため、前培養の段階及び接合までをLB培地に変更し、若干の改善は見られたが、大半の実験は生育の観測または定量化が出来なかった。これは供与菌側の染色体DNAの接合移動後の相同組換えの効率低下が大きな要因の可能性と考え、相同組換えの時間を長く取ることを検討した。

そこで1次スクリーニングは供与菌の排除、2次スクリーニングは受容菌の排除と組換え体の選択、その後3次スクリーニングで組換え体のみを選択という工程での検討を行なった。その結果、2次スクリーニングをCmのみで受容菌を排除する選択にした場合、多くの実験でコロニーの定量化に成功した。全体で3回の選択を行う事で、染色体DNAの接合移動後の培養時間が伸び、相同組換えのチャンスが増えることで2重欠失を獲得した株の数が増えた事が、成功率の上昇の要因の一つと考えられる。2次スクリーニングでKm・Cm両方を用いてセレクションを行い、さらに3次スクリーニングも同様にKm・Cmの二重選択を続けた場合に改善は見られなかった。これは二重薬剤選択のストレスによって、相同組換え効率の低下によりDKO株の数の減少に起因するものと考えられた。2次スクリーニングをCmのみで行うことで大腸菌にかかるストレスが低下した影響で組換え効率が維持され、より多くのDKO株を得る事が出来た可能性が考えられる。これらの結果を踏まえ、前培養及び接合はLB培地で、1次スクリーニングはKm、2次スクリーニングはCm、そして3次スクリーニングはKm・Cmの二重選択で組換えによる二重遺伝子欠失株の選択を、最少培地によるDKO株作製標準プロトコルとした。

また、確立した最少培地を用いたDKO株作製プロトコルを用いた実験はその殆どが定量化を行うことができ、また再現性を確認した結果十分に高いことがわかった。しかし、継代培養の回数を増やした事で大腸菌の生育時間が長くなっていることから染色体上に欠失した遺伝子のコピーであるPartial Duplication (PD) や生育環境に適応するような変異の挿入が起こる確率が上がっている事が予想出来る。今回はこれらの確認は行っていないため、3度の継代培養がDKO株の染色体構造に与える影響に関しては検証を行う必要があると考えている。検証は全体のうちどれだけの株にPDが発生しているか、欠失が保持されているかを確認する事で行うことが出来る。当研究室ではKeio collectionで欠失されている全ての遺伝子に対して、その野生型遺伝子の同定

を行うプライマーを保有している。そのため、全体から見てどれだけの割合でPDや変異が発生しているかはPCRを行う事で確認出来ると考えられる。

3 章

遺传的相互作用解析結果を用いた探索による補完経路を担う酵素遺伝子の推定

3.1 緒言

2章では4種の糖源を炭素源とする最少培地でのシミュレーション実験の予測生育値と生物実験の実測値とが異なる遺伝子欠失株に着目しDKO株観測実験を行おうと試みた。予備実験では殆どが観測出来ないという結果になったが、プロトコルの改変によって観測および解析が出来るようになった。確立した標準プロトコルを用いてDKO株生育観測実験を行うことでFNを示す遺伝子の受容菌とKeio collectionを掛け合わせた網羅的なDKO株の生育度データの取得を行った。その結果、264個の実験条件でMGRの値を取得する事が出来た。1実験では1プレート内のコロニー数1536個のデータを得る事ができ、大腸菌は約4288のタンパク質をコードする遺伝子が報告されており、Keio collectionはそのうち3864遺伝子の欠失株を網羅するので、1セットの生育観測データは1遺伝子につき3実験3プレートに相当する。よって、一つの受容菌に対する網羅的なDKO株生育観測は3実験3プレートで一つの単位となる。2章で得られた264実験の生育度観測データを受容菌ごとに3実験で一纏めに出来る場合のみを抽出した所、68遺伝子に関するデータを得る事が出来た。本章では、この網羅的な68x4000=272000個のデータの中から頑強性に寄与する代替経路を担う未知の遺伝子を見つけ出すために、FNを示す遺伝子に対し遺伝的相互作用を示す遺伝子の検出及び解析を行った。遺伝的相互作用の検出には一般的には ϵ スコア(Collins et al., 2006; Tomanek et al., 2020)が使用される。 ϵ スコアとは、序章にて述べたalleviateかaggravateかを判断する数値の事であり、SKO株の生育度より計算したDKO株の予想生育度から実際のDKO株の生育度を引くことで計算される。二つのSKO株の生育度をそれぞれ ΔA , ΔB としDKO株の生育度を ΔAB とすると ϵ は以下の式になる。

$$\epsilon = \Delta AB - \Delta A \cdot \Delta B$$

本研究では、DKO株生育観測実験で得られた生育度全てに対して ϵ スコアを計算し、負の遺伝的相互作用を示す遺伝子の検出を試みた。

まず、フィルタリング及び正規化の手法を検討しDKO株の生育データに含まれるノイズを取り除くようにデータクレンジングを行った。その後、処理されたデータを

用いて ϵ スコアを計算し、FNを示す遺伝子に対して負の遺伝的相互作用を持つ遺伝子を網羅的に検出した。検出された遺伝子の中にはFNを示す遺伝子を代替する機能を担うものがあると予想されるため、既知の情報と照らし合わせ現在の代謝ネットワークモデルに記載されていない代替機能の推定を行った。また、過去に酵母のチームから、DKO株生育観測実験の生育度プロファイルを用いたクラスタリングによる機能未知遺伝子の機能推定が報告されている(Costanzo et al., 2016)。そのため、本研究でもそれに倣い今回のDKO株生育観測実験から得たMGRのプロファイルを用いてクラスタリングを行い同様の機能推定が行えるか検証を行った。以上の検証による結果を踏まえ、本研究の目的である代謝ネットワークモデルを用いたシミュレーション実験においてFNの原因とされる未知の代替経路を担う遺伝子の同定を試みる。

3.2 手法

3.2.1 生育のばらつきと落ち込みによるデータの抽出

生育度スコアのMGRを用いてX軸を遺伝子の染色体上の位置、Y軸をMGRとしてXY平面上にプロットし遺伝子のゲノム上の位置でMGRをプロットし可視化したものをゲノムマップと呼ぶ。今回得られたデータ全てのゲノムマップを描くと、ゲノムマップ全体のMGRが大きくばらついているデータとばらつきが小さいデータ(図24)、特徴的なMGRの落ち込みが発生しているデータ(図26A)が存在していたため、これらの評価を行い解析に使用するデータの選別を行った。まず、MGRの落ち込みは染色体上の2箇所に見られ、1箇所目は受容菌側で欠失されている遺伝子の位置周辺、2箇所目はHfr化に使用したCIPが挿入されている領域周辺であると予想出来る。そのため、この2箇所付近の遺伝子が示す生育度の集合とその他の領域の集合を比較し、違いを検出することが出来たならば落ち込みが確認出来たと言えろと考え、2群間の検定を行うこととした。MGRの落ち込みが予想される遺伝子は、1箇所目は受容菌側で欠失される遺伝子から上下300kbpの範囲内に位置しているものを対象とした。また、2箇所目はHfr化に使用したCIPが挿入されている箇所から150kbpまでの範囲内に位置している遺伝子を対象とした。これらの領域とその他の領域の違いを検出するために用いた検定は2群間の平均値が独立であり正規分布が仮定できないためWilcoxonの順位和検定とする事とした。Wilcoxonの順位和検定はPythonのScipy1.5.1モジュールにあるstats関数を使用し行った。

次にゲノムマップ全体のMGRのばらつきの評価方法については、ばらつきの大きさを判別するために定量化する必要がある。そのため、全データに対してそれぞれゲノムマップ全体のMGRの分散値を求めばらつきの指標として使用した。大きくばらついているデータの検出には、分散値から頻度分布を描き閾値を決める事によって行おうと考えた。分散値はPython3.8.0とnumpy1.19.0モジュールにあるvar関数を用いて計算した。

3.2.2 欠損値の補填方法

2章で取得した網羅的なDKO株の生育度データには遺伝子欠失による生育の落ち込みと相同組み換え効率の低下による見かけの生育の落ち込みの2つの要因が混在している。よって、後者が要因の影響を取り除き遺伝子欠失による生育の落ち込みのみを評価出来るようにするため、補正を行う必要があると考えた。補正に関しては、まず補正したい遺伝子とゲノム上において前後5遺伝子の計11遺伝子の生育度から中央値を計算し、その中央値とゲノム全体の中央値の差を補正したい遺伝子に足しあわせて行った。この補正を全遺伝子の生育度に対して行う事で、相同組み換え効率の低下による生育の落ち込みを補正した。また、過去当研究室において供与菌株による遺伝子欠失の挿入箇所から±30kbpとCIPの挿入箇所から30kbpの範囲では相同組み換え効率の低下による影響が大きくDKO株が作製できないと言われている(Typas et al., 2008a)。よって、その範囲内にある遺伝子のデータは全て除外することとした。しかし、除外したことでデータに欠損値が発生してしまい、その後の解析において障害となってしまう。そのため、データ全体の平均値±標準偏差の範囲からランダムな値を生成し欠損値の補填を行った。この手法は欠損値の補填によく使われており、データ全体の分布の形に影響を与えることなく欠損値を補填する事が可能である(Wes McKinney., 2017)。

3.2.3 εスコアの計算

生育度の落ち込みとばらつき、そして移動中央値を用いた補正と欠損値の補填を行った後、処理されたデータに対して遺伝的相互作用の検出のためにεスコアの計算を行った。

$$\varepsilon = \Delta AB - \Delta A \cdot \Delta B$$

予測値の計算には2章にて取得したM9 Glucose培地でのSKO株の生育度を使用した。その後、統計処理を行うためにεスコアのスケールを揃えるためZ scoreを計算し有意に負の遺伝的相互作用を示す遺伝子を検出した。

3.2.4 生育プロファイルに基づいたクラスタリング

受容菌株側の網羅的なゲノムマップに基づいたクラスタリングを行うことで機能別に別れたクラスターを得ることが出来るという酵母のチームからの報告(Costanzo et al., 2016)を参考に、本研究で得られた網羅的なDKO株の生育データから得たゲノムマップを用いてクラスタリングを行い、機能別に別れたクラスターを得る事が可能か検証を行う事とした。本研究では供与菌株は代謝経路上の酵素を担う遺伝子の欠失株を使用しているため、受容菌株で欠失されている遺伝子も同様に酵素を担うもののみに限定しクラスタリングを行った。酵素かどうかを分類する際にはKEGG Release 83.0のEC番号情報を使用した。遺伝子を分割するクラスターの数はギャップ統計量から導出した。ギャップ統計量の計算にはR 4.0.2のclusGap関数を使用した。クラスタリング手法にはK-means法を指定し、サンプル数が大きいためモンテカルロ法のブートストラップ回数は標準的な手法に従い100回に設定しギャップ統計量を計算した。次に提案されたクラスター数で酵素に限定した受容菌株側の生育度プロファイルをクラスタリングし、クラスター内に機能の集合が起こっているか確認した。機能別分類にはCOGs id (2014版)を使用した。

3.2.5 負の遺伝的相互作用検出

ϵ スコアをまとめたテーブルに対してターゲットとなる供与菌、そして受容菌であるKeio collectionそれぞれから見てZ score化を行った。Z scoreの計算にはscipyのZ score関数を使用した。Z scoreは-2を閾値とし、供与菌と受容菌どちらから見ても閾値以下となっている組み合わせを抽出し、負の遺伝的相互作用を示す遺伝子の組み合わせを抽出した。この中から機能が類似している遺伝子がペアになっていないかどうか確認するため、配列が類似しているかどうかに着目しblastpを用いて配列類似性を計算した(Altschul et al., 1990)。

3.3 結果

3.3.1 生育のばらつきの定量化および評価

2章で得られた網羅的なDKO株生育データには、画像データからの観測時のエラーや測定条件に起因するノイズが多数含まれている。そのため、解析を行う前に異常値の処理やデータの補正、またデータの確からしさの定量化を行い、データクレンジングを前処理として行う必要があった。本研究で行うべき前処理を検討するため、得られたそれぞれの生育度データの確認を行った所、特徴的な傾向が二つあることがわかった。まず、ゲノムマップを確認すると、1つ目の特徴は、 $y=1$ から大きく離れた生育度を示す株が多数存在するばらつきの大きいデータ(図24A)がある事である。殆どのDKO株はwtと同様に生育するため $y=1$ の直線上周辺にプロットされる(図24B)。2つ目の特徴は、供与菌株によって挿入される欠失遺伝子の付近とHfr株に挿入されているCIP領域付近に位置する遺伝子の欠失株が、低い生育度を示す傾向がある事である(図26A)。これらに対してどの特徴を持つデータを使用すべきなのか判断するため検討を行った。

まずデータ内でばらつきの大きいゲノムマップであった実験と、ばらつきの小さいゲノムマップであった実験に関しては生育曲線を描くと、ばらつきの大きかった方は小さかった方と比較してLag timeが長い事を確認した。例として供与菌株に*moaA*のHfr株を供与菌として使用した2実験のゲノムマップと生育曲線を示す(図24)。同一の遺伝子欠失株を使用した実験であってもデータのばらつきが大きいとLag timeが長いことが図24から確認出来るため、欠失した遺伝子による影響が原因ではないことがわかる。接合を用いたDKO株作製においてLag timeが長くなる現象は過去に当研究室で確認されており、Lag timeが長い場合はDKO株になっていない大腸菌が薬剤耐性遺伝子を長い時間をかけて獲得してしまい適応した結果生育していたと議論されている[竹内博士論文]。また、接合の際に受容菌の前培養の状況によって接合頻度が大幅に変わる事がわかっており、今回行った二重遺伝子欠失実験では前培養の状態を確認しながら実験を行ったが、一部の実験で前培養が上手くいっておらず接合頻度の低下が起これ二重遺伝子欠失株が上手く得られずデータがバラついてしまった事も考えられる。よって、ゲノムマップのばらつきが大きい実験のデータはDKO株の生育観測データが正しく得られていない可能性を考慮し解析の対象から除

外する必要がある。そのため、何らかのデータ処理を行いばらつきの小さいデータのみを全体から選別する必要があると考えた。

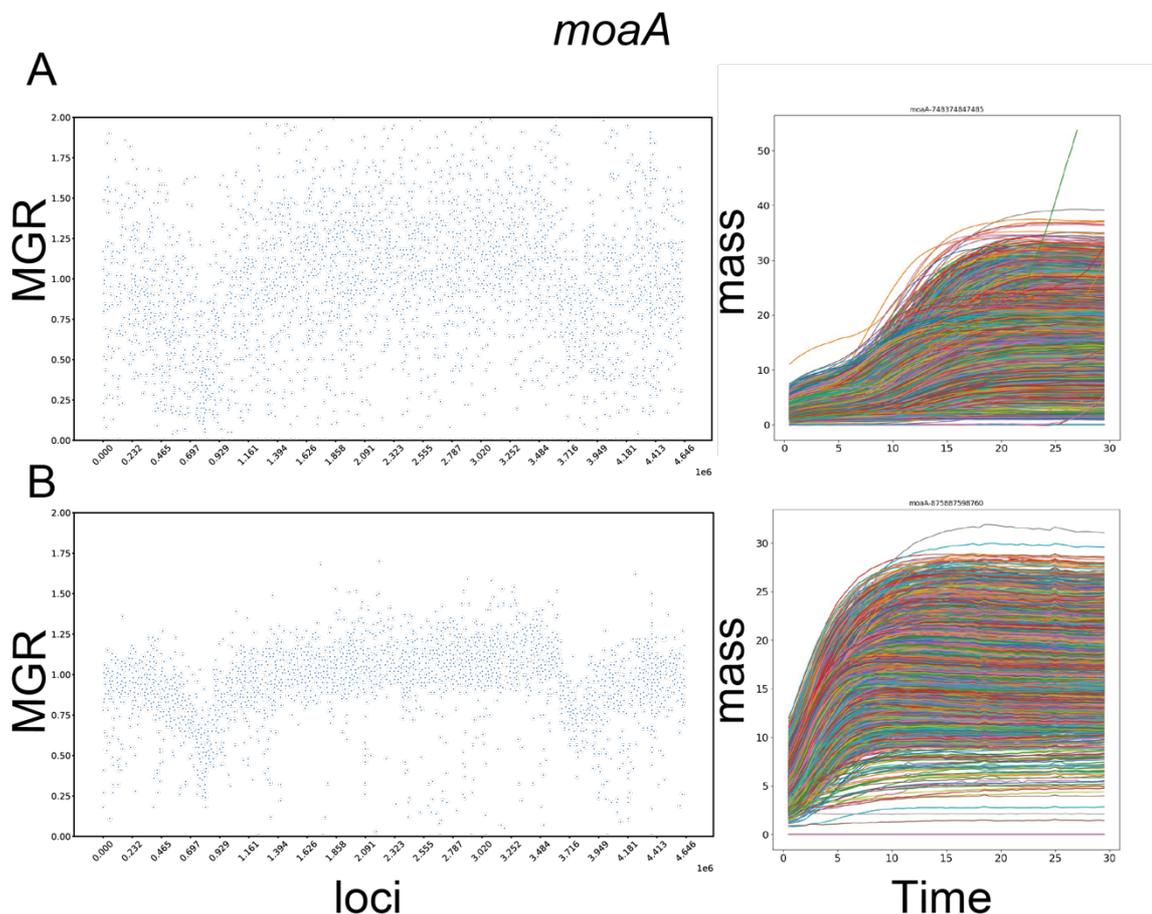


図24. データのばらつきと生育曲線の関係

X軸を遺伝子の染色体上の位置、Y軸を生育スコアMGRとしてプロットするとデータ全体のばらつきが大きいデータ(A左)と小さいデータ (B左)があった。また、ばらつきの大きいデータの生育曲線(A右)を確認するとLag timeがばらつきの小さいデータ(B右)に比べて長かった。

ゲノムマップ全体のMGRのばらつきの検出には、ばらつき度合いを定量化する必要がある。そのため1実験における全てのMGRから分散値を計算し、全てのDKO株生育観測における分散値の頻度分布を描くことでばらつきの大きいデータと小さいデータを分ける閾値を求めることが出来ると考えた。図25に、全てのDKO株生育観測データからそれぞれ求めた分散値の頻度分布を示す。図25より、DKO株の生育度データは0.15を閾値とすることでばらつきの大きいデータと小さいデータに分ける事が出来ると考え、分散が0.15以下のデータのみを抽出した。結果、68遺伝子の内

47遺伝子の生育データに関してばらつきが少ない事がわかった。

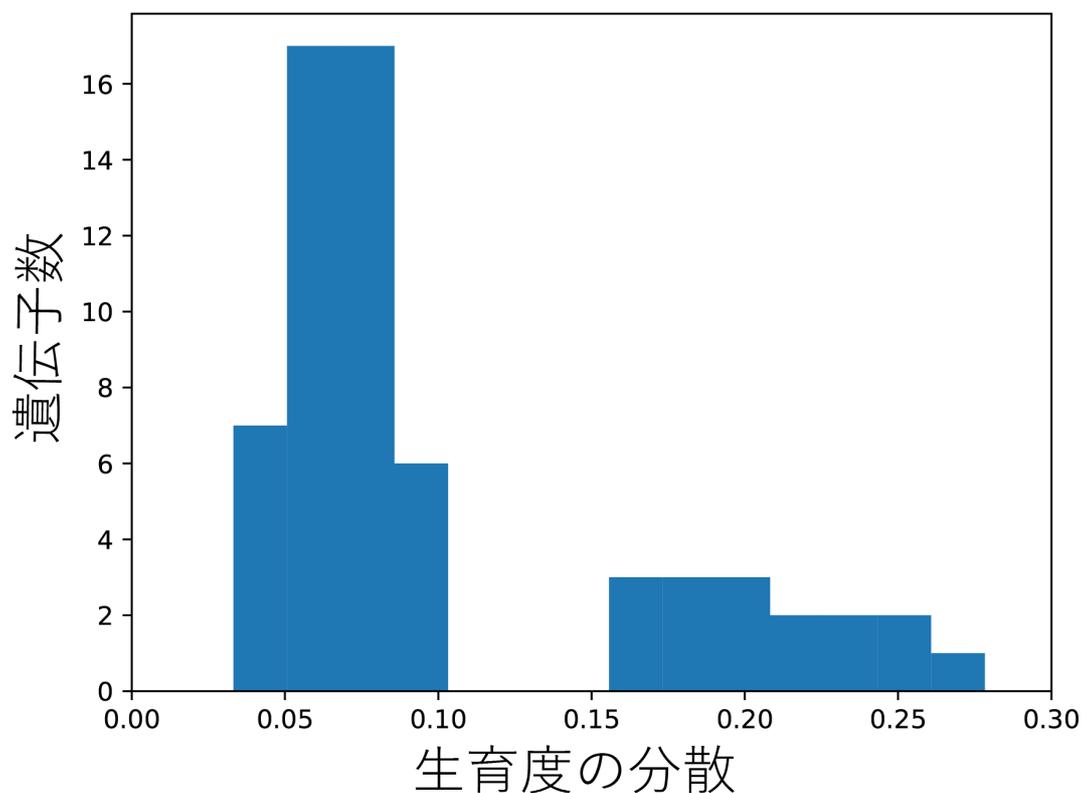


図25. DKO株の網羅的な生育データから計算したそれぞれの実験での分散の頻度分布

3.3.2 生育の落ち込みの原因と検出および評価

3.2.1で述べたように、ゲノムマップを描くとMGRの急激な落ち込みが発生している箇所を確認する事が出来る。この落ち込みが予想される領域とその他の領域の2群に分け、その間に違いが生じているかWilcoxonの順位和検定を用いて判別できるか確認を行うこととした。MGRの落ち込みが発生しているデータ (A) と発生していないデータ (B) を図26に示す。

まず、供与菌株によって挿入される欠失遺伝子の付近で急激に生育が落ち込んでいる現象に関してはLB培地を用いた実験において既に過去に報告されている(Cote et al., 2016)。しかし、M9培地を用いた実験における、Hfr化に使用したCIPが挿入されている領域付近の落ち込みについての報告は確認出来なかった。

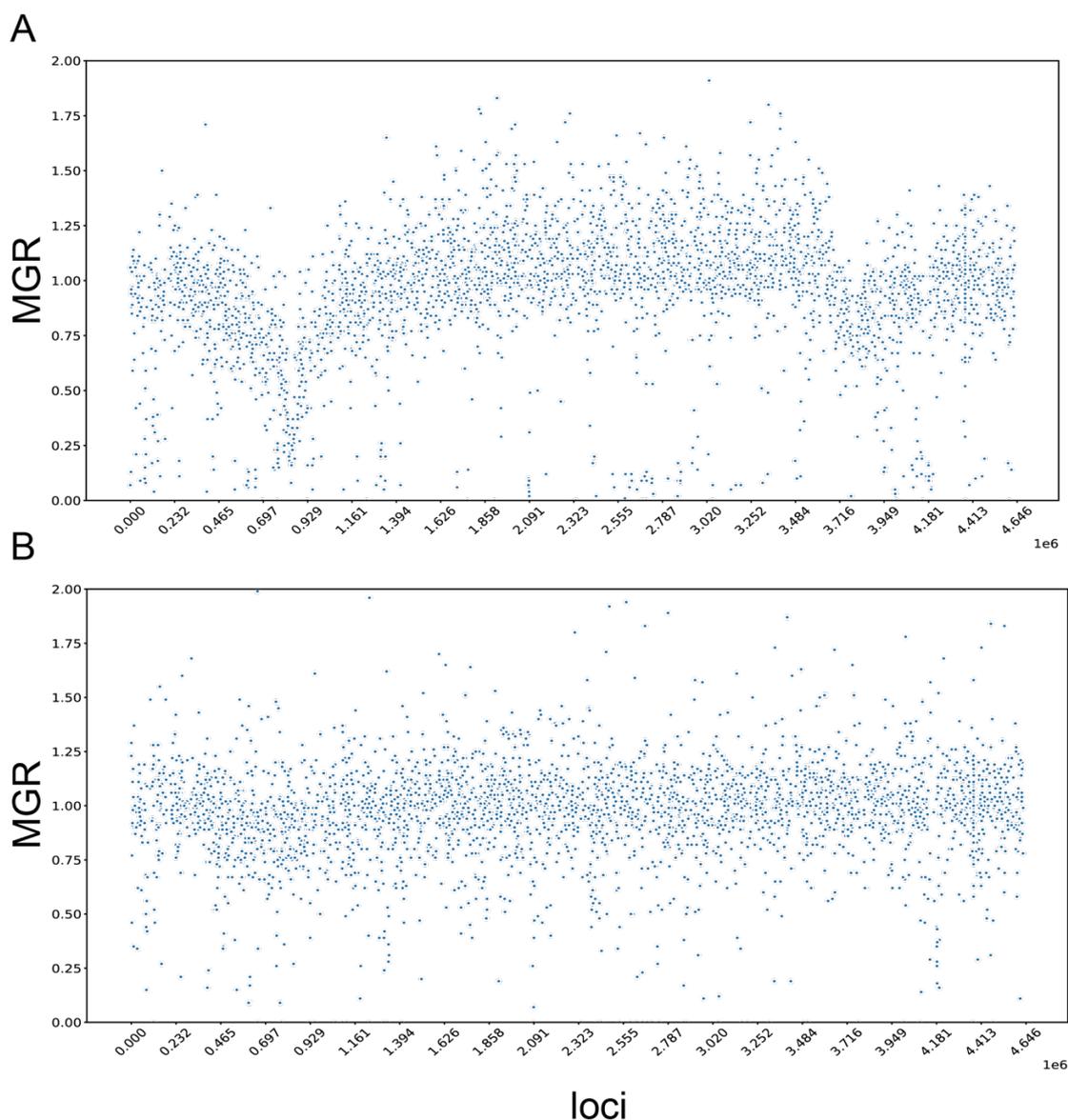


図26. 各DKO株の生育速度の分布と染色体上の位置

欠失遺伝子の染色体上の位置とその二重欠失株の生育。A) CIPの挿入部位で生育の落ち込みが発生しているデータ。B) 生育速度の落ち込みが見られないケース。

何故供与菌株によって挿入される欠失遺伝子の付近に急激な生育の落ち込みが起こるのかに関しては、相同組み換えの効率の低下が原因である。目的のDKO株の2つの欠失遺伝子の染色体上の距離（図27中のD）が小さい時、相同組み換えを起こしうる範囲が狭くなるため、結果として相同組み換えが起こる確率が小さくなる。つまり相同組み換え効率が低下する事でDKO株となった大腸菌の数が少なくなってしまう、見かけ上の生育の落ち込みが発生してしまう。

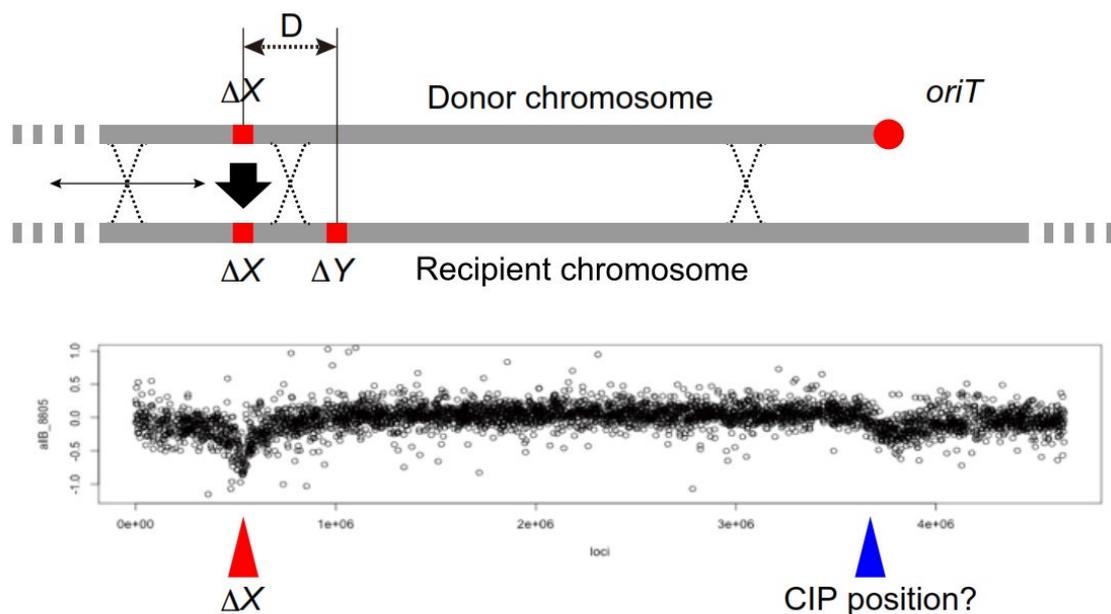


図27. DIPの発生機構

ΔXは供与菌側の遺伝子欠失、ΔYは受容菌側の遺伝子欠失。上段上部のバーは接合移動により受容菌側に移動してきた供与菌染色体DNAを、上下段のバーは受容菌側の染色体DNAを表す。相同組換えで供与菌側のΔXが相同組換えにより受容菌側染色体に移るためには、ΔXとΔY及びΔXの左側の間の組換えが必要になる。下段は受容菌側欠失遺伝子の染色体上の位置とその生育速度(MGR)を描いたゲノムマップである。

次に、Hfr化に使用したCIPの挿入箇所が落ち込む場合であるが、これは接合の際に複製起点 $oriT$ からシングルストランドの染色体が複製され、受容菌に接合移動するが、長時間の接合を行なっているので、染色体1周以上が受容菌側に移動するものと考えられる。その場合、図28に示す様に、CIPが挿入されている両側を持つDNAが存在するようになると考えられる。その場合、 $oriT$ 近傍の領域が特に相同組換え頻度が高い場合、CIP挿入近傍のΔYが抜け落ちる確率が高まることから、CIP近傍のΔYとのDKO株のMGRが落ち込む可能性がありそうだが、そもそも $oriT$ 近傍が相同組換え頻度に違いを産む機構が存在するかどうか、検討が必要であろう。

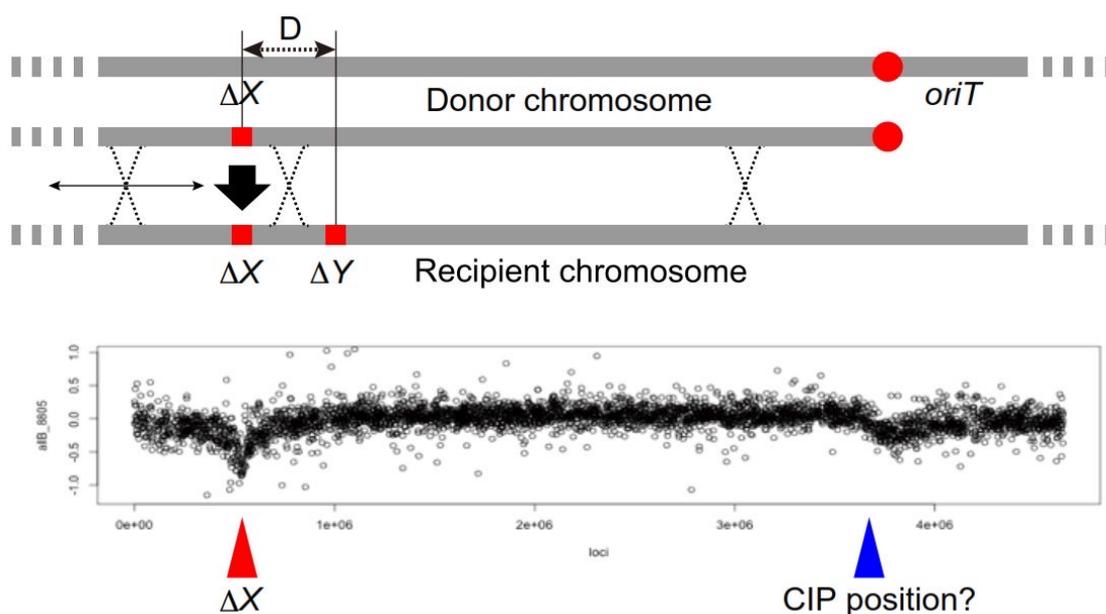


図28. Second DIPの発生機構

供与菌側染色体DNAが、全染色体1周以上移動した場合の受容菌側の状態

以上より、DKO株の作製に接合を用いているため供与菌株側で欠失されている遺伝子の位置周辺とHfr化に使用したCIPが挿入されている箇所周辺には見かけ上生育の落ち込みが発生することから、落ち込みが発生しているデータを補正する必要があると考えた。よって、急激な生育の落ち込みを、落ち込みが予想される領域とその他の領域の2群に分け、2群の間に違いが生じているかWilcoxonの順位和検定を用いて検出できるか確認を行った。なお、生育の落ち込みが発生すると思われる範囲内のデータ数は約400、それ以外の範囲のデータ数は約3000であった。全てのデータのp値に対して \log_{10} を計算した値の頻度分布を図29に示す。図29より、 $\log_{10}p=-20$ からデータが大きく別れていることがわかったため、 $\log_{10}p$ が-20より小さい値となったデータのみを抽出した。その結果、43遺伝子に対応するデータが $\log_{10}p<-20$ を示した。

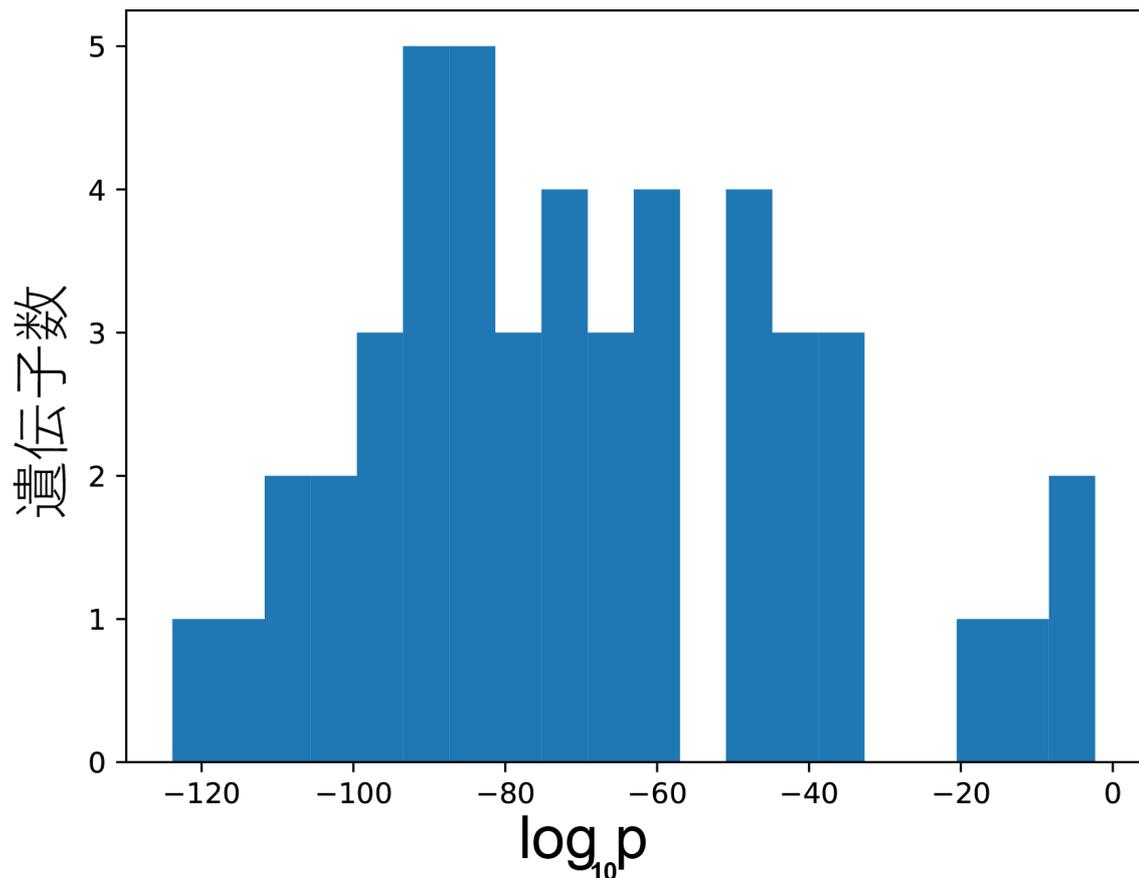


図29. Wilcoxon順位和検定によるDIPの同定

DIPが発生して生育が落ち込んでいる範囲の遺伝子群とそれ以外の範囲に違いがあるかを定量化するため行ったWilcoxonの順位和検定によって得られたp値の頻度分布

以上より、ばらつきと生育の落ち込みの2つの指標からフィルタリングを行った結果、2章で得られた網羅的なDKO株の生育データのうち43遺伝子のデータはばらつきが小さく、かつ見かけの生育の落ち込みが2箇所発生している事がわかった。今後のデータ処理と解析は43遺伝子(受容菌側の遺伝子の重複を除くと37遺伝子)のデータに対してのみ行うこととした。概要を表13に示す。

表13. 各データ処理によってフィルタリングを通った実験数及び供与菌として利用した遺伝子欠失株数

	実験数	遺伝子ごとに纏めた数	遺伝子の重複を除いた数
標準化に成功した数	264		
Keio collection 3プレートがそろっている数	204	68	51
分散によるフィルタリングを通過した数	141	47	41
検定によるフィルタリングを通過した数	129	43	37

3.3.3 相同組み換え出来ない箇所の除外と補填方法

ばらつきと生育の落ち込みの2つの指標からフィルタリングされた43個のデータには、遺伝子欠失による生育の落ち込みと先に述べた相同組み換え効率の低下による生育の落ち込みの2つの要因が混在している。そのため、このままでは遺伝的相互作用の検出に使用出来ないため、相同組み換え効率の低下による生育の落ち込みを補正する必要がある。相同組み換え効率による生育の落ち込みは受容菌株欠失遺伝子の染色体上の位置に依存するため、過去の報告からゲノムポジションに沿った移動中央値を用いた補正が提案されており(Cote et al., 2016)、本研究でも同様に補正に移動中央値を用いた。補正の流れとしては、まず補正したい遺伝子とゲノム上において前後5遺伝子の計11遺伝子のMGRから中央値を計算し、中央値とゲノム全体のMGRの中央値の差を補正したい遺伝子に足しあわせるといった手順で行った。この補正を全遺伝子の生育度に対して行う事で、相同組み換え効率の低下による生育の落ち込みを除いた。また、過去当研究室において供与菌株による遺伝子欠失の挿入箇所から $\pm 30\text{kbp}$ とCIPの挿入箇所から 30kbp の範囲では相同組み換え効率の低下による影響が大きくDKO株が作製できない事が確認されている(Typas et al., 2008a)。よって、その範囲内にある遺伝子のデータは全て除外することとした。しかし、除外したことでデータに欠損値が発生してしまい、その後の解析において障害となってしまう。この状態はMissing At Randomと呼ばれるパターンに近いと見られるため、多重代入法としてデータ全体の平均値 \pm 標準偏差の範囲からランダムな値を生成し欠損値の補填を行った。この手法はMissing At Randomな欠損値の標準的な補填法の一つであり、データ全体の分布の形

に影響を与えることなく欠損値を除外する事が可能である。

3.3.4 接合を用いた DKO 実験における遺伝的相互作用スコア ϵ

ゲノム全体の生育度のばらつきや落ち込みを定量化し選別を行い、移動中央値や欠損値が補填されデータクレンジングを済ませた網羅的なDKO株生育データを用いて、遺伝的相互作用スコアである ϵ の計算を行った。その後 ϵ スコアをZ score化し、ヒートマップを描くことで負の遺伝的相互作用を示す遺伝子の検出を試みた。しかし、ヒートマップを描くと供与菌株に*mhpT*欠失株を使用している場合ほとんどの遺伝子が負の遺伝的相互作用を示している（図30Aヒートマップの左赤色部分）事と、供与菌株に*gltA*欠失株を使用している場合ほとんどの遺伝子が正の遺伝的相互作用（図30Aヒートマップの左緑色部分）を示す事がわかった。そのため、*mhpT*と*gltA*について確認を行ったところ、M9 Glucose上でWTの生育度MGRを1とした時に*mhpT*の欠失株はMGRが1.42と非常に高く、*gltA*の欠失株はMGRが0.17と非常に低い事が分かった。Colony-liveではプレート内のコロニーの総massが等しいという前提のもと定量値が標準化されている。SKOであれば、プレート上のコロニーは異なる位置に遺伝子欠失を持っているため、プレート上のコロニー生育度の平均はWTの生育度に近似できる。しかし、本研究で用いた網羅的なDKO株の構築法ではプレート上の全てのコロニーが供与菌株側の欠失を保有することになり、プレート上のコロニー生育度の平均は供与菌株の生育度に近似される。このことからColony-liveで標準化したDKO株の生育度定量値をそのまま上記の ϵ 計算式に代入することは出来ない。そのため、本研究ではColony-liveで標準化した定量値に使用できるように遺伝的相互作用スコアを再定義することとした。計算式は ΔAB の代わりに、DKO株生育度の観測値 ΔAB に供与菌株の生育度 ΔA を掛けることで求めたDKO株の生育度の推定値 Δab を用いるように再定義した。本研究ではこの遺伝的相互作用スコアを ϵ_2 と呼ぶこととする。 ϵ_2 の計算式は以下の通りである。

$$\epsilon_2 = \Delta ab - \Delta A \cdot \Delta B \text{ (但し } \Delta ab = \Delta A \times \Delta AB \text{)}$$

この ϵ_2 の計算式を用いて、網羅的なDKO株生育データから ϵ_2 を計算しZ score化しヒ

ートマップを描いて(図31B)確認したところ、殆どの遺伝子に対して負の遺伝的相互作用を示していた*mhpT*が他の供与菌株と同様のプロファイルの挙動を示した。*gltA*に関しては、SKO株の生育度が低すぎるため、殆どの値が0付近になるという結果になった。しかし ϵ スコアを用いた場合に見られた、図31Aに見られるような*gltA*が殆どの遺伝子に対して正の遺伝的相互作用を示した傾向は、図31Bでは失われている事から網羅的なデータからの遺伝的相互作用の検出という目的においては改善されている事がわかる。

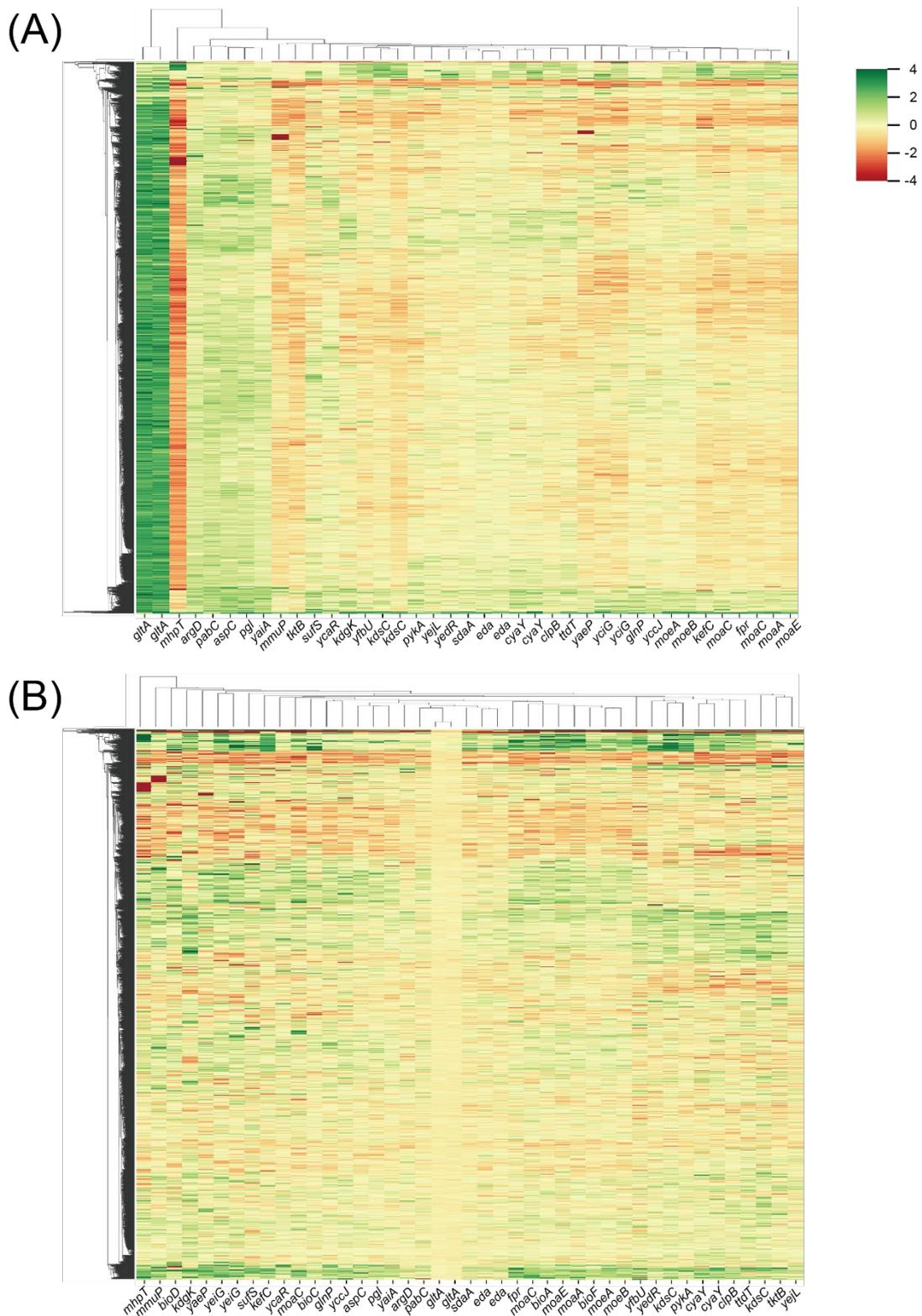


図30. (A) ϵ スコアを用いて計算したZ scoreのヒートマップ、(B)再定義した ϵ スコアを用いて計算したZ scoreのヒートマップ

3.3.5 生育度プロファイルに基づいたクラスタリング

本研究で得られたDKO株の網羅的な生育データから ϵ スコアを計算し、受容菌側からみた ϵ スコアの変動、つまりそれぞれの遺伝子が示す ϵ プロファイルの特徴からクラスタリングを行い、クラスター内に機能が集合するか、そして機能未知遺伝子の機能予測が出来るか検証を行った。今回使用した供与菌株で欠失している遺伝子は全て代謝関連であるため、クラスタリングに使用する受容菌株側の生育プロファイルは酵素を担う遺伝子のみに限定した。酵素であるかの判定はKEGGデータベースを参照しEC番号が割りふられているかで行った。その結果826の酵素を担う遺伝子の生育度プロファイルをクラスタリングに使用した。クラスター数はRの`clustGap`関数を用いて計算したギャップ統計量から最適な値を求めた結果、7クラスターに分割した。それぞれのクラスターの生育プロファイルをヒートマップとして図31に示す。図31より、クラスター2とクラスター5が他のクラスターと比べて正の遺伝的相互作用を示す組み合わせが多く、特に*mhpT*とクラスター5に分類された遺伝子は強く正の遺伝的相互作用を示すという特徴から独立したクラスターとして検出されたと考えられる。

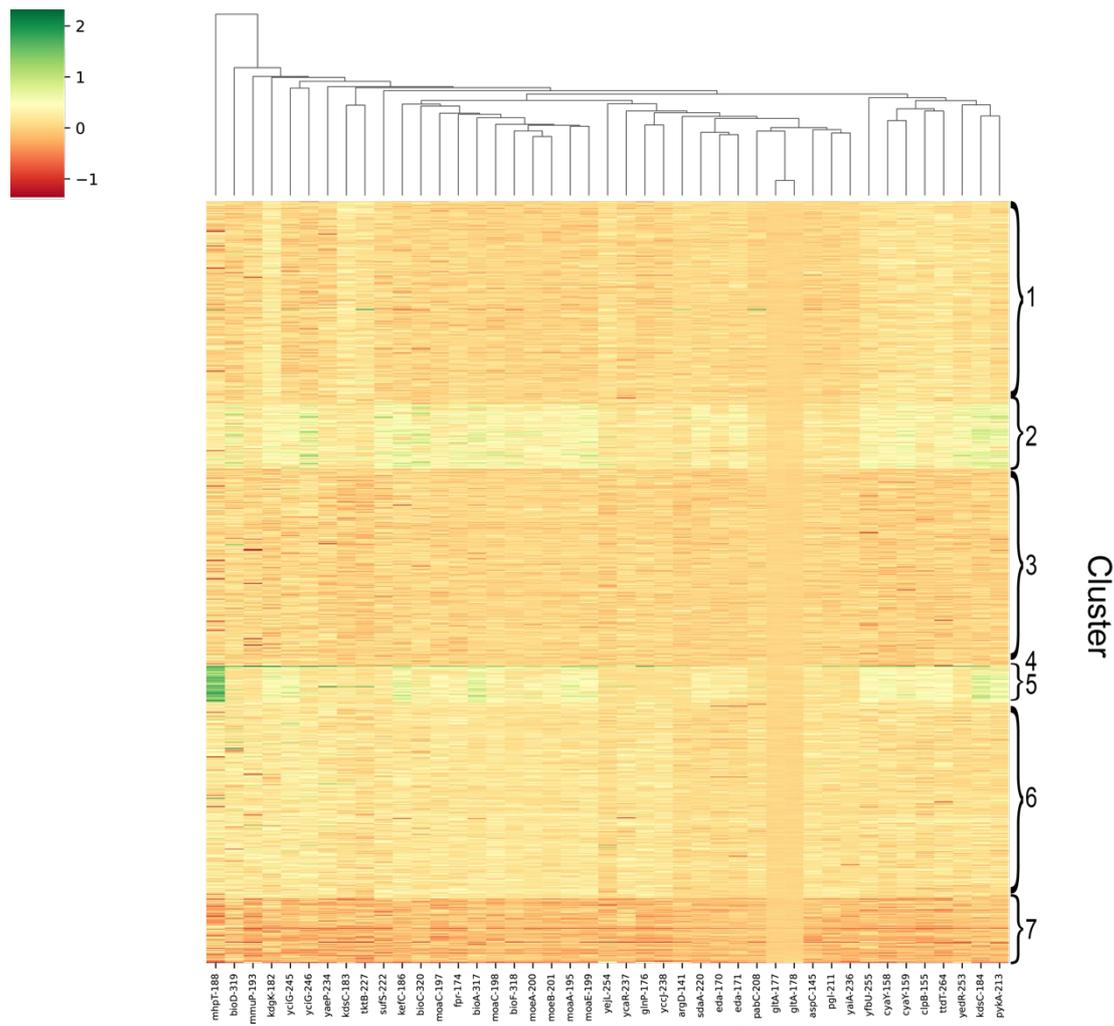


図31. 7つのクラスターの生育プロファイルを描いたヒートマップ

クラスター2と5に分類された遺伝子は全体的に生育が良く、それ以外は全体的に生育が低いという特徴があった。

次に、COGsを用いてそれぞれのクラスターに分類された遺伝子の機能を調べた。結果を図32に示す。図32より、先程全体的に正の遺伝的相互作用を示すクラスター2とクラスター5にヌクレオチドとアミノ酸代謝関連の遺伝子の機能の偏りが確認出来た。特にクラスター5は*mhpT*と組み合わせた際に生育が良くなるという特徴があり、そのためクラスターになっていた。しかし、両方のクラスター内に機能未知の遺伝子は含まれていなかった。また、機能未知を示す遺伝子が含まれていたクラスター1, 3, 6に関しては、ヌクレオチド代謝やリン酸代謝に若干の機能の偏りは見られるもののクラスター内の遺伝子が満遍なく様々な機能に分類されており大きな機能

の偏りは確認できず、目的であった機能未知遺伝子の機能推定を行うことは出来なかった。

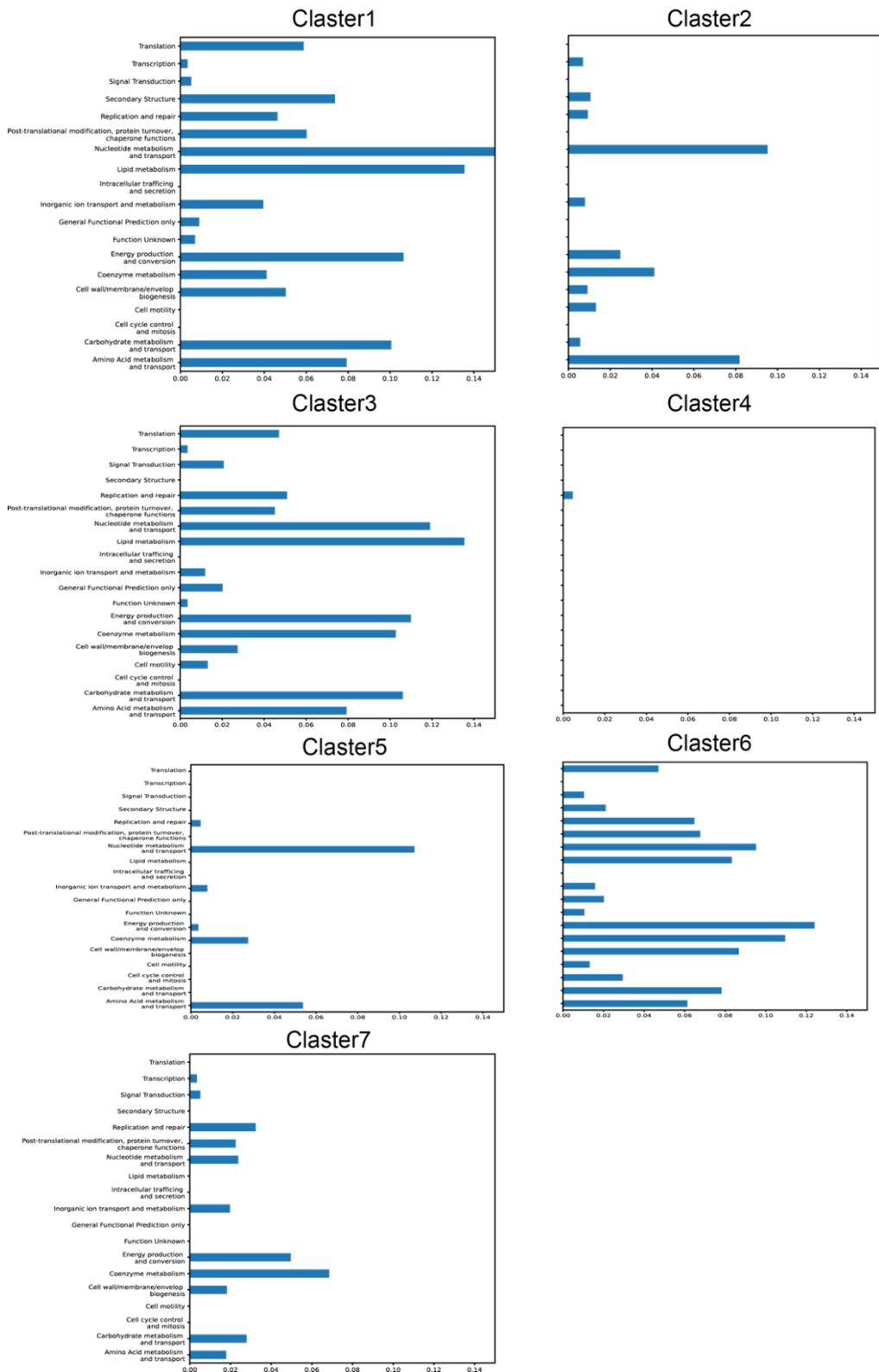


図32. COGs情報に基づいて分類した7つのクラスター内それぞれの遺伝子機能

X軸はCOGsごとの遺伝子の総数でクラスターに分類された内のCOGsごとの遺伝子数を割る事で計算したCOGsごとの割合である。Y軸はCOGsの分類を示す。クラスター2と5にはヌクレオチドとアミノ酸代謝関連の遺伝子が多く分類された。

3.3.6 False negative を示す遺伝子と配列類似性および負の遺伝的相互作用を示す遺伝子

本章にて得られた $\epsilon 2$ スコアのテーブルを受容菌側、供与菌側から見た場合ごとにZ score化し、双方で $Z \leq -2$ となった1178通りの遺伝子の組み合わせを抽出し、これを負の遺伝的相互作用を示す遺伝子ペアとした。この1178通りの組み合わせの中で、配列が似ている組み合わせがないかBlastPを用いて検索したところ、表14に示す6つの組み合わせで配列の相同性があることがわかった。

表14. 負の遺伝的相互作用を示した1178ペアのうち、配列が類似していた6ペアの組み合わせ

受容菌側の欠失 遺伝子	供与菌側の欠失 遺伝子	Z score_row	Z score_col	E-Value
<i>argD</i>	<i>gabT</i>	-2.76	-2.96	2.00E-74
<i>aspC</i>	<i>tyrB</i>	-3.57	-4.22	2.00E-113
<i>kdgK</i>	<i>frlD</i>	-3.65	-2.39	7.00E-07
<i>mhpT</i>	<i>yaaU</i>	-5.75	-3.24	1.00E-07
<i>pykA</i>	<i>pykF</i>	-2.61	-2.09	7.00E-89
<i>tkkB</i>	<i>tktA</i>	-3.6	-4.63	0.00E+00

このうちピルビン酸キナーゼである*pykA*、*pykF*とトランスケトラーゼである*tkkB*、*tktA*に関してはM9 Glucose上でFNではなく、isozymeとして既知の関係であった。残りの4ペアに関しては現在までisozymeであるという報告はされておらず、これらを対象に解析を進めることとした。

まずそれぞれの遺伝子にアノテーションされた情報を確認した結果、ArgDはN-Acetylornithine、GabTはbeta-Alanineのアミノ酸基を2-Oxoglutarateが反応に寄与してケトン基に転移させていた(図33A)。また、AspCはL-Aspartate、TyrBはL-Phenylalanineのケトン基を2-Oxoglutarateが反応に寄与してアミノ酸基に転移させていた(図33B)。KdgKは2-Dehydro-3-deoxy-D-gluconate、FrlDはFructoselysineのヒドロキシ基をATPによってリン酸基にしていた(図33C)。また、化学反応だけでなく、それぞれの基質の構造も図33より類似している事が分かり、よって基質特異性が類似していると予想出来る。図33にそれぞれの遺伝子が担う酵素の化学反応式を示す。

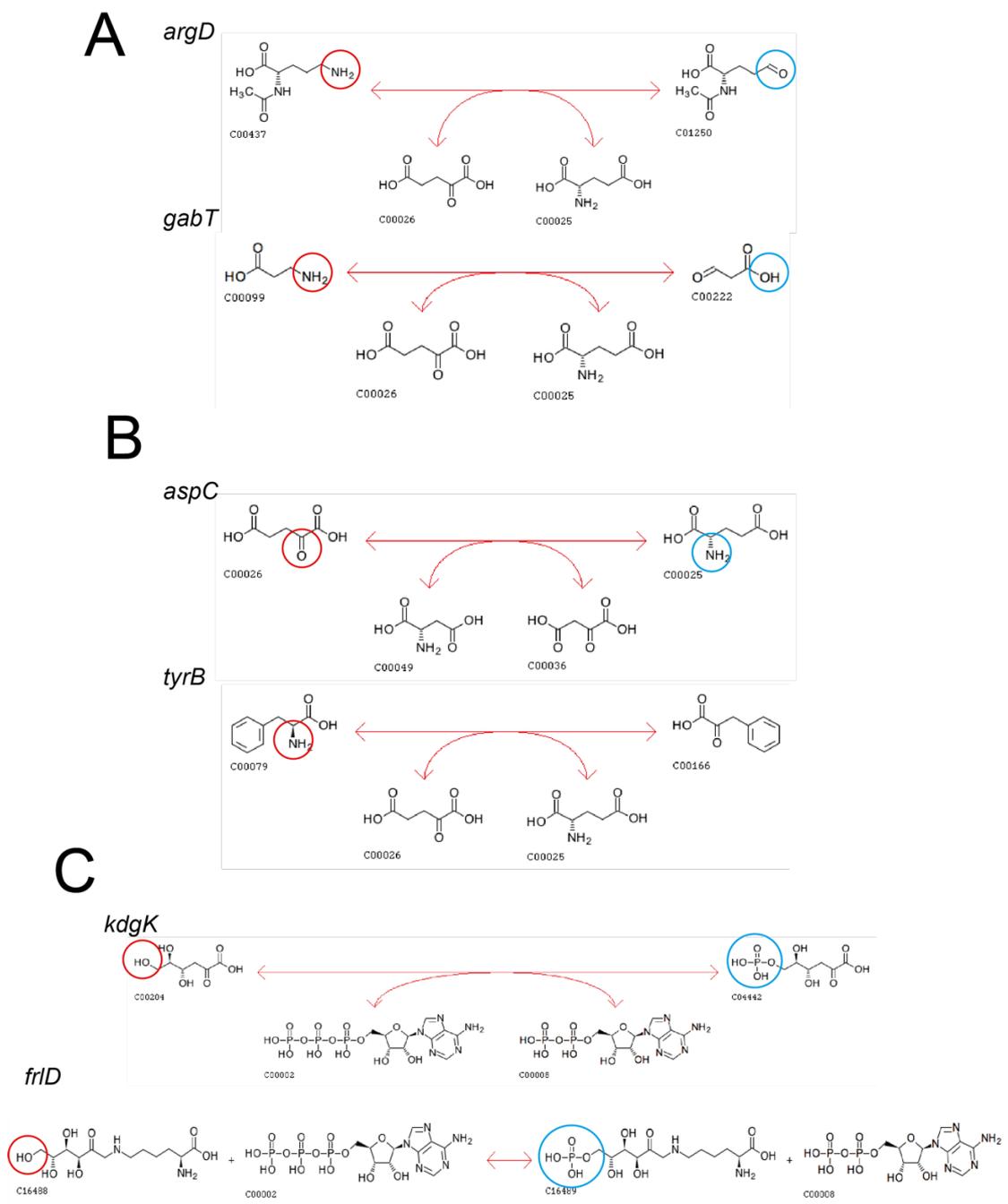


図33. 反応機構の相同性

A) ArgDとGabTそれぞれの酵素反応 (B) AspCとTyrBそれぞれの酵素反応 (C) KdgKとFrlDそれぞれの酵素反応。どの場合も基質の構造が似通っており、反応の対象となっている部位と反応の種類が同じである。

3.4 考察

本章でデータ処理によってフィルタリングされた43個のDKO株生育データから $\epsilon 2$ を計算し、クラスタリングによる機能の集合の確認と負の遺伝的相互作用を示す遺伝子からFNを示す遺伝子を代替する機能を担う因子の推定を行った。

クラスタリングは、 $\epsilon 2$ のプロファイルに基づき酵素を担う遺伝子に限定して行った。結果、2つのクラスターに大きな機能の集合を確認したがその他のクラスターに関しては大きな機能の偏りは確認出来なかった。また、ヌクレオチドとアミノ酸代謝関連の遺伝子が集合したクラスター2, 4には機能未知もしくは機能が推定されている遺伝子が含まれていなかったが、クラスター1, 3, 6には機能が分かっていない遺伝子が複数含まれていた。この3つのクラスターは他のクラスターに比べて分類された遺伝子の数が3倍近く多く、そのためどの機能にも満遍なく遺伝子が分類されてしまったと考えられる。よって、クラスター1, 3, 6を対象にさらなるクラスタリングを行うことで、機能未知の遺伝子がどの機能集団に集合するか確認し機能推定が可能である可能性が考えられる。また、機能未知に分類されている遺伝子群に機能の偏りが無いという場合今回のクラスタリングではデータ数が約40遺伝子分と少ない事から大きく機能が集合しなかった可能性も考えられる。

次に、 $\epsilon 2$ 全体をRowまたはColumn方向にZ score化しどちらからみても $z \leq -2$ を示す遺伝子を、FNを示した遺伝子と負の遺伝的相互作用を示す遺伝子としてどのように代替機能を担うか推定を行った。代替機能にisozymeを想定し、FNを示した遺伝子と負の遺伝的相互作用を示す遺伝子の配列の相同性をBlastpによって確認を行った。その結果、既にisozymeとして報告されているものが2例、そして既知の情報からは関係性が報告されていないものが4例あることがわかった。このうち、ArgDとGabTは2-Oxoglutarateを用いてアミノ酸基をケトン基に、AspCとTyrBは2-Oxoglutarateを用いてケトン基をアミノ酸基に、KdgKとFrlDはATPを用いてヒドロキシ基をリン酸基にするといったように同様の反応を行っており、また基質の構造も似通っている事がわかった。そのため、これら3ペアに関してはisozymeの関係である可能性が高いと考えている。また、これら3つの遺伝子ペアが何故似たような基質に対して同じような反応を行なっているかに関しては、祖先となる遺伝子が重複を起こし、そこ

から基質特異性を獲得した結果2つの遺伝子となったためであると推察出来る。また *mhpT* と相同性を示した *yaaU* はトランスポーターと機能が予測されている遺伝子であるが (Pao et al., 1998)、それ以外の5つが全て構造の似ている基質に共通の反応を行っていることから、*yaaU* も *mhpT* と同じ機能を有するトランスポーターの遺伝子であると考えられる。MhpTは3-ヒドロキシフェニルプロピオン酸トランスポーターであり (Ferrández et al., 1997)、3-ヒドロキシフェニルプロピオン酸の取り込みを担うプロトンポンプとして働いている (図34)。

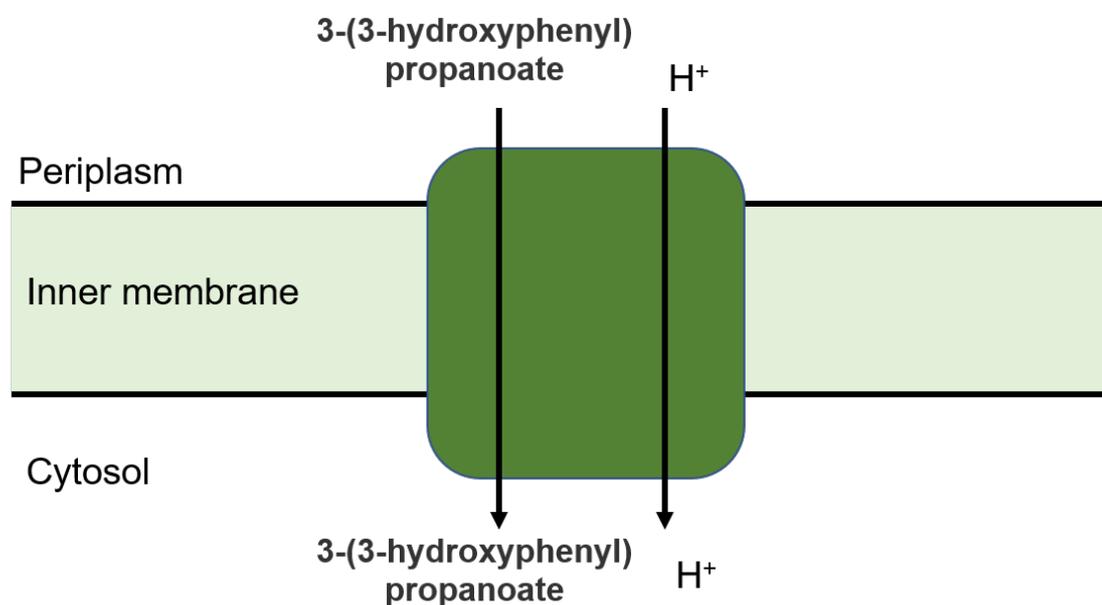


図34. MhpTは3-ヒドロキシフェニルプロピオン酸の取り込みに関わるプロトンポンプである

MhpTと負の遺伝的相互作用を示しかつ配列が似ていることから、YaaUも3-ヒドロキシフェニルプロピオン酸トランスポーターであるプロトンポンプである可能性が高いことが予測できる。以上より網羅的なDKO株の生育度観測実験から負の遺伝的相互作用を計算することで、FNを示す遺伝子 *argD*, *aspC*, *kgdK* に対してそれぞれ配列類似性を示す *gabT*, *tyrB*, *frlD* が得られ、それらは未報告のアイソザイムとして機能を代替していることによってFNの原因となっている可能性が示唆された。また、機能がほとんどわかっていない遺伝子 *yaaU* と *mhpT* の類似性が示され、代謝経路上の未知の因子の推定を行うことが出来たと考えられる。

文献

- Arifuzzaman, M., Maeda, M., Itoh, A., Nishikata, K., Takita, C., Saito, R., Ara, T., Nakahigashi, K., Huang, H.C., Hirai, A., *et al.* (2006). Large-scale identification of protein-protein interaction of *Escherichia coli* K-12. *Genome Res* 16, 686-691.
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2, 2006 0008.
- Babu, M., Arnold, R., Bundalovic-Torma, C., Gagarinova, A., Wong, K.S., Kumar, A., Stewart, G., Samanfar, B., Aoki, H., Wagih, O., *et al.* (2014). Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in *Escherichia coli*. *PLoS Genet* 10, e1004120.
- Baryshnikova, A., Costanzo, M., Kim, Y., Ding, H., Koh, J., Toufighi, K., Youn, J.Y., Ou, J., San Luis, B.J., Bandyopadhyay, S., *et al.* (2010). Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nature methods* 7, 1017-1024.
- Benjamin Gompertz. (1825). On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. *Philosophical Transactions of the Royal Society of London* Vol. 115 , pp. 513-583.
- Bernhard Ø. Palsson. (2015). *Systems Biology -Constraint-based Reconstruction and Anlysis* - (Cmbridge University Press).
- Blattner, F.R., Plunkett, G., 3rd, Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F., *et al.* (1997). The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453-1474.
- Boogerd, F.C., Boe, L., Michelsen, O., and Jensen, P.R. (1998). *atp* mutants of *Escherichia coli* fail to grow on succinate due to a transport deficiency. *Journal of Bacteriology* 180, 5855-5859.
- Burgard, A. P., Nikolaev, E. V., Schilling, C. H., & Maranas, C. D. (2004). Flux coupling analysis of genome-scale metabolic network reconstructions. *Genome research*, 14(2), 301–312.
- Butland, G., Babu, M., Diaz-Mejia, J.J., Bohdana, F., Phanse, S., Gold, B., Yang, W., Li, J., Gagarinova, A.G., Pogoutse, O., *et al.* (2008). eSGA: *E. coli* synthetic genetic array analysis. *Nat Methods* 5, 789-795.
- Christie, P.J. (2001). Type IV secretion: intercellular transfer of macromolecules by systems ancestrally related to conjugation machines. *Mol Microbiol* 40, 294-305.
- Collins, S.R., Schuldiner, M., Krogan, N.J., and Weissman, J.S. (2006). A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol* 7, R63.
- Costanzo, M., VanderSluis, B., Koch, E.N., Baryshnikova, A., Pons, C., Tan, G., Wang, W., Usaj, M., Hanchard, J., Lee, S.D., *et al.* (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353.

- Cote, J.P., French, S., Gehrke, S.S., MacNair, C.R., Mangat, C.S., Bharat, A., and Brown, E.D. (2016). The Genome-Wide Interaction Network of Nutrient Stress Genes in *Escherichia coli*. *MBio* 7.
- Dulbecco, R. (1986). A turning point in cancer research: sequencing the human genome. *Science* 231, 1055-1056.
- Edwards, J., Ibarra, R. & Palsson, B. (2001). In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat Biotechnol* 19, 125–130.
- F., R. (1962). A new puffing pattern induced by temperature shock and DNP in *Drosophila*. *Experientia* 18, 571-573.
- Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., and Palsson, B.O. (2007). A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* 3, 121.
- Ferrández, A., Garcíá, J.L., and Díaz, E. (1997). Genetic characterization and expression in heterologous hosts of the 3-(3-hydroxyphenyl)propionate catabolic pathway of *Escherichia coli* K-12. *J Bacteriol* 179, 2573-2581.
- Fowler ZL, Gikandi WW, Koffas MAG (2009) Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production. *Appl Environ Microbiol* 75: 5831–5839
- Fuhrer T, Chen L, Sauer U, Vitkup D (2007) Computational prediction and experimental verification of the gene encoding the NAD⁺ / NADP⁺-dependent succinate semialdehyde dehydrogenase in *Escherichia coli*. *J Bacteriol* 189: 8073–8078
- Galperin, M.Y., Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2015). Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res* 43, D261-269.
- H Kitano. (2001). *Foundations of Systems Biology*. MIT Press, Cambridge, Mass., London, p.297.
- Hu, B., Khara, P., and Christie, P.J. (2019). Structural bases for F plasmid conjugation and F pilus biogenesis in *Escherichia coli*. *Proc Natl Acad Sci U S A* 116, 14222-14227.
- Kaizu, Kazunari, Nishida, Kozo, Sakamoto, Yuki, Kato, Suguru, Niina, Toru, Nishida, Naoki, Aota, Naohiro, Koizumi, Moriyoshi, & Takahashi, Koichi. (2020). E-Cell System version 4 (v2.1.0b2). Zenodo.
- Kanehisa, M. and Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* 29, 28-35.
- Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M.V., Jacobs, J.M., Bolival, B., Jr., Assad-Garcia, N., Glass, J.I., and Covert, M.W. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389-401.
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2006). Complete set of ORF clones of *Escherichia coli* ASKA library (A

- Complete Set of E. coli K-12 ORF Archive): Unique Resources for Biological Research. *DNA Research* 12, 291-299.
- Klobucar, S.L., Gaeta, J.W., and Budy, P. (2018). A changing menu in a changing climate: Using experimental and long-term data to predict invertebrate prey biomass and availability in lakes of arctic Alaska. *Freshwater Biology* 63, 1352-1364.
- Klopfenstein, D.V., Zhang, L., Pedersen, B.S. et al. (2018). GOATOOLS: A Python library for Gene Ontology analyses. *Sci Rep* 8, 10872.
- K. Roberts, P. Walter, Michael R. Green, J. Sambrook. (2015). *Molecular Cloning A Laboratory Manual Fourth Edition*. Engineering
- Lederberg, J., Cavalli, L.L., and Lederberg, E.M. (1952). Sex Compatibility in *Escherichia Coli*. *Genetics* 37, 720-730.
- Lederberg, J., and Tatum, E.L. (1946). Gene recombination in *Escherichia coli*. *Nature* 158, 558.
- Majewski, R.A., and Domach, M.M. (1990). Simple constrained-optimization view of acetate overflow in *E. coli*.
- Monk, J.M., Lloyd, C.J., Brunk, E., Mih, N., Sastry, A., King, Z., Takeuchi, R., Nomura, W., Zhang, Z., Mori, H., et al. (2017). iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nature biotechnology* 35, 904-908.
- Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H., and Tomita, M. (2009). Systematic phenome analysis of *Escherichia coli* multiple-knockout mutants reveals hidden reactions in central carbon metabolism. *Mol Syst Biol* 5, 306.
- O., P.B. (2015). *Systems Biology - Constraint-based Reconstruction and Analysis*.
- Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B.O. (2011). A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism-2011. *Mol Syst Biol* 7, 535.
- Orth, J.D., and Palsson, B. (2012). Gap-filling analysis of the iJO1366 *Escherichia coli* metabolic network reconstruction for discovery of metabolic functions. *BMC Syst Biol* 6, 30.
- Orth, J.D., Thiele, I., and Palsson, B.O. (2010). What is flux balance analysis? *Nature biotechnology* 28, 245-248.
- Oshima, T., Aiba, H., Masuda, Y., Kanaya, S., Sugiura, M., Wanner, B.L., Mori, H., and Mizuno, T. (2002a). Transcriptome analysis of all two-component regulatory system mutants of *Escherichia coli* K-12. *Mol Microbiol* 46, 281-291.
- Oshima, T., Wada, C., Kawagoe, Y., Ara, T., Maeda, M., Masuda, Y., Hiraga, S., and Mori, H. (2002b). Genome-wide analysis of deoxyadenosine methyltransferase-mediated control of gene expression in *Escherichia coli*. *Mol Microbiol* 45, 673-695.
- Otsuka, Y., Muto, A., Takeuchi, R., Okada, C., Ishikawa, M., Nakamura, K., Yamamoto, N., Dose, H., Nakahigashi, K., Tanishima, S., et al. (2015). GenoBase: comprehensive resource database of *Escherichia coli* K-12. *Nucleic Acids Res* 43, D606-617.

- Pal C, Papp B, Lercher MJ (2005) Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat Genet* 37: 1372–1375.
- Pal C, Papp B, Lercher MJ, Csermely P, Oliver SG, Hurst LD (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* 440: 667–670.
- Pao, S.S., Paulsen, I.T., and Saier, M.H., Jr. (1998). Major facilitator superfamily. *Microbiol Mol Biol Rev* 62, 1-34.
- P Novick, B C Osmond and D Botstein. (1989). Suppressors of yeast actin mutations. *GENETICS* April 1, 1989 vol. 121 no. 4 659-674.
- Reed, J.L., Vo, T.D., Schilling, C.H., and Palsson, B.O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol* 4, R54.
- Rutherford, S.L., and Lindquist, S. (1998). Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336-342.
- Takeshi, N., Agata, Tamaki, Ai, M., Yuichiro, T., and Hirotada, M. described elsewhere.
- Takeuchi, R., Tamura, T., Nakayashiki, T., Tanaka, Y., Muto, A., Wanner, B.L., and Mori, H. (2014). Colony-live -a high-throughput method for measuring microbial colony growth kinetics- reveals diverse growth effects of gene knockouts in *Escherichia coli*. *BMC Microbiol* 14, 171.
- Taylor, A.L., and Thoman, M.S. (1964). THE GENETIC MAP OF *ESCHERICHIA COLI* K-12. *Genetics* 50, 659-677.
- Tomanek, I., Grah, R., Lagator, M., Andersson, A.M.C., Bollback, J.P., Tkacik, G., and Guet, C.C. (2020). Gene amplification as a form of population-level gene expression regulation. *Nat Ecol Evol* 4, 612-625.
- Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., *et al.* (2008a). A tool-kit for high-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nature Methods* 5, 781-787.
- Typas, A., Nichols, R.J., Siegele, D.A., Shales, M., Collins, S.R., Lim, B., Braberg, H., Yamamoto, N., Takeuchi, R., Wanner, B.L., *et al.* (2008b). High-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nat Methods* 5, 781-787.
- Varma A, Palsson BO. (1993). Metabolic capabilities of *Escherichia coli*: I. synthesis of biosynthetic precursors and cofactors. *J Theor Biol.* 1993 Dec 21;165(4):477-502.
- Wes McKinney. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd. ed.). O'Reilly Media, Inc.
- Yamamoto, Y., Aiba, H., Baba, T., Hayashi, K., Inada, T., Isono, K., Itoh, T., Kimura, S., Kitagawa, M., Makino, K., *et al.* (1997). Construction of a contiguous 874-kb sequence of the *Escherichia coli* K-12 genome corresponding to the 50.0-68.8 min on the linkage map and analysis of its sequence features (supplement). *DNA Res* 4, 169-178.
- Zhao, D., Kong, Y., Zhao, S. et al. (2020). Engineering Functional DNA–Protein Conjugates for Biosensing, Biomedical, and Nanoassembly Applications. *Top Curr Chem (Z)* 378, 41.