

令和 2 年 5 月 15 日現在

機関番号：12501

研究種目：若手研究(B)

研究期間：2017～2019

課題番号：17K15769

研究課題名(和文) 侵襲性無莢膜型インフルエンザ菌の全ゲノム解析による感染メカニズムの解明

研究課題名(英文) The analysis of whole genome sequences in invasive nontypeable Haemophilus influenzae strains

研究代表者

真下 陽一 (Mashimo, Yoichi)

千葉大学・大学院医学研究院・技術専門職員

研究者番号：90422253

交付決定額(研究期間全体)：(直接経費) 3,200,000円

研究成果の概要(和文)：本研究は、侵襲性感染症を引き起こすインフルエンザ菌無莢膜株(NTHi)に注目し、侵襲性に関わる因子を同定するために次世代シーケンサーを用いて全ゲノム配列の決定を行った。ロングリードの次世代シーケンサー(PacBio)を用いることにより、NTHi株の全ゲノム配列を高精度に決定することができ、また、リファレンスゲノム配列との比較により各株の変異を同定することができた。今後、さらに全ゲノム配列を収集し、解析を進めることで侵襲性に関わる因子などの同定を目指す。

研究成果の学術的意義や社会的意義

Hibワクチンの普及以降、NTHi株による侵襲性感染症の報告が増加しており、病原菌の侵襲性に関与する因子には未解明の部分が多く存在している。また、薬剤耐性菌の増加も問題であり、その対策が重要となっている。対策としてまず、原因となる起炎菌について正確に知ることが挙げられ、正確な全ゲノム配列を決定することが診断法、治療薬、ワクチンの開発などに有用となると期待される。

研究成果の概要(英文)：In this study, we determined the whole genome sequences of nontypeable Haemophilus influenzae (NTHi) strains to identify invasive factor(s) in invasive NTHi infections. Highly accurate whole genome sequences of NTHi strains were obtained using long-read next-generation sequencing platforms (PacBio) and mutations in each strain were detected by comparison with reference sequences published in NCBI as complete genomes. In future studies, we plan to collect more whole genome sequences of H. influenzae strains to identify important factors related to, such as invasiveness, pathogenicity, toxicity and antimicrobial susceptibility.

研究分野：ゲノム医科学

キーワード：インフルエンザ菌 侵襲性感染症 無莢膜株 全ゲノムシーケンス 次世代シーケンサー

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

(1) 侵襲性感染症は、血液や髄液など無菌的な部位から起炎菌が分離される場合の総称であるが、上気道に定着した細菌が、ウイルス感染症などをきっかけとして血液中に侵入した後、髄液、中枢神経系に感染することで発症する。小児におけるその主な原因菌は、インフルエンザ菌と肺炎球菌である。細菌性髄膜炎などの重症感染症は、早期診断が難しいこと、抗菌薬に対する耐性菌が増加していること、治療に成功しても聴力障害や麻痺などの重度の後遺症を残すことが多いから、予防をはじめとする様々な対策が求められる。

(2) インフルエンザ菌 (*Haemophilus influenzae*; Hi) には莢膜株と無莢膜株 (NTHi) があり、莢膜株のうち最も病原性の高い b 型 (Hib) に対する侵襲性感染症を予防するため、現在は Hib ワクチンが定期接種化されている。インフルエンザ菌の場合、Hib の持つ b 型の莢膜が侵襲性に重要であると考えられていたが、Hib ワクチンの普及以降は莢膜を持たない NTHi による侵襲性感染症も報告されており、病原菌の侵襲性に関与する因子に未解明の部分も多く存在している。

(3) 細菌のゲノム変化は極めて早く、他の細菌との間で遺伝子水平伝播も頻繁に起こっていることも明らかになっており、未知の機能を持ったゲノム領域が新たに生じ、侵襲性や薬剤耐性などに影響を及ぼすことも十分に考えられる。このため、それらを明らかにするためには精度の高い全ゲノム配列情報が極めて重要となる。

2. 研究の目的

本研究では、侵襲性感染症を引き起こす NTHi 株を中心に、ロングリードの次世代シーケンサーを用いて正確な全ゲノム配列を決定し、比較解析することで、感染メカニズムの解明に重要な侵襲性に関与する因子や薬剤耐性に関与する因子など感染症対策に有用な領域を同定することを目指す。

3. 研究の方法

(1) NTHi 試料 DNA

日本国内において 2008 年～2018 年の期間に侵襲性感染症を引き起こした患者の血液または髄液から単離し保存していた、NTHi 株 28 株をチョコレート寒天培地にて培養し、ゲノム DNA の抽出を行った。また、非侵襲性感染症症例として肺炎患者の喀痰から分離された NTHi 株 10 株、健康保菌者の上咽頭より採取・分離された NTHi 株 10 株についても同様にゲノム DNA の抽出を行った。

(2) ロングリードの次世代シーケンサー (PacBio) によるゲノム配列データの取得

抽出したゲノム DNA に対して、g-TUBE (コパリス社) を用いて断片長が 8~12kbp となるように断片化し、アダプター付加等を行い、SMRTbell ライブラリを調製した。調製した各ライブラリを次世代シーケンサー PacBio RS II または Sequel (パシフィック・バイオサイエンス社) で解析し、ゲノム配列データを取得した。得られた配列データに対して、パシフィック・バイオサイエンス社が提供する HGAP ソフトウェアを用いた de novo assemble を行い、コンティグ配列を決定した。

(3) ショートリードの次世代シーケンサー (Illumina) によるゲノム配列データの取得

抽出したゲノム DNA に対して、Nextera DNA サンプル調製キット (イルミナ社) を用いて断片化されたライブラリを調製した。調製した各ライブラリを MiSeq 試薬キット v2 (500 cycles、イルミナ社) を用いて次世代シーケンサー Illumina MiSeq (イルミナ社) で解析し、ペアエンドのゲノム配列データを取得した。得られた配列データに対して、Platanus v1.2.4 ソフトウェアを用いた de novo assemble を行い、コンティグ配列を決定した。

(4) データ解析

PacBio と Illumina の両方の次世代シーケンサーでゲノム配列データを取得した株については、PacBio で構築されたコンティグ配列と Illumina のペアエンドの配列データを利用して、Pilon v1.22 ソフトウェアを用いた assembly polishing を行い、ハイブリッドコンティグ配列を得た。このハイブリッドコンティグは、相同性の高い領域や繰り返し配列を正確に決定できるというロングリードの長所と、塩基あたりの精度が高いというショートリードの長所を併せたものであるため、より正確なゲノム配列になっていると考えられる。MUMmer ソフトウェアを用いて、ハイブリッドコンティグと、PacBio および Illumina で構築されたコンティグ配列を比較することにより mismatches 部位およびギャップ部位の探索を行い、PacBio および Illumina で構築されたコンティグ配列の特性について検討した。

次に、公共データベースである NCBI において完全長ゲノムが公開されているインフルエンザ菌 14 株の配列情報をリファレンス配列として、決定した NTHi 株の全ゲノム配列データとの比較を BLAST の相同性検索を利用した独自のプログラムを用いて行うことで、配列一致率および変異を決定した。

またインフルエンザ菌の生物型は、ウレアーゼ活性、オルニチンデカルボキシラーゼ活性、インドール産生の 3 種類の酵素活性の有無により I~ の 8 種類に分類されるが、決定した NTHi 株の全ゲノム配列データを用いてこれらの酵素活性に関連する遺伝子、*ureA-ureH*、*speF*、*tnaA* における変異について調べ、生物型との関連を確認した。

4. 研究成果

(1) ロングリードの次世代シーケンサー (PacBio) による全ゲノム配列データの取得

収集した NTHi 株 48 株について、次世代シーケンサー PacBio RS II または Sequel を用いてゲノム配列データを取得し、HGAP を用いた de novo assemble を行ったところ、それぞれにおいて配列長 1.79 ~ 1.98Mbp の単一のコンティグ配列が構築された。それぞれのコンティグ配列は両端の約 10kb の配列が一致し、環状のゲノム構造であることが確認できた。さらに、それぞれの株において読み取り深度は >100 であり、データの信頼度を示す QV 値は >48 (塩基が正しくない確率がおよそ 0.001%以下) であったため、高精度な全ゲノム配列が得られたと言える。また、単離後の培養中に菌株がどの程度の新たな変異を獲得するかどうかを評価するために、NTHi 株 1 株を用いて継代を行う前後での全ゲノム配列の比較も行った。継代の前後での配列の一致率は 99.9997% であり、違いは全て 7bp 以下リピート配列の繰り返し回数であった。わずかに変異が生じているものの大きなゲノム構造の変化は認められなかったことから、短期間の培養中に新たな変異を獲得する確率は低く、解析には影響しないと考えられた。

(2) ショートリードの次世代シーケンサー (Illumina) によるゲノム配列データとの比較

収集した NTHi 株 48 株のうち 11 株については、ショートリードの次世代シーケンサー Illumina MiSeq を用いてゲノム配列データを取得した。Platanus v1.2.4 を用いて de novo assemble を行った結果、24 ~ 83 本のコンティグ配列が構築され、合計配列長は 1.75 ~ 1.94Mbp であった。対応する PacBio で構築されたコンティグ配列と比較すると、Illumina で構築されたコンティグ配列の合計配列長は 29.8 ~ 48.0kbp 短くなっており、Illumina では構築できなかった領域が存在していると考えられた。

さらに、PacBio で構築されたコンティグ配列と Illumina のペアエンドの配列データを利用してハイブリッドコンティグを構築し、PacBio のコンティグ配列との比較を行った結果、配列長に 0 ~ 25bp の差があり、0 ~ 29 か所見られたミスマッチ部位はほとんどが同じ塩基が連続する homopolymer の塩基数の違いであった。

一方、ハイブリッドコンティグと Illumina で構築されたコンティグ配列との比較を行った結果では、4 ~ 10 か所見られたミスマッチ部位は大部分が繰り返し配列によるものであり、一部の株のデータでは相同性の高い領域に misassemble していると考えられるものも見られた。加えて、19 ~ 55 か所、全長にして 29.8 ~ 48.0kbp のギャップ部位も見られ、その一番の要因は相同性の高い領域の存在であり、次いで繰り返し配列の存在も要因であると考えられた。

PacBio による全ゲノム配列の決定では、homopolymer の配列は正確に決定しにくいものの、相同性の高い領域や繰り返し配列も正確に決定できることが確認できた。

(3) リファレンス配列との比較による変異の同定

公共データベース NCBI において完全長ゲノム配列が公開されているインフルエンザ菌 14 株の配列情報をリファレンス配列として、全ゲノム配列を決定した NTHi 株 20 株とを一对比較した結果、塩基配列レベルでの一致率が 71 ~ 91% であった。NTHi 株間においても一致率は 70 ~ 92% であり、かなりの多様性が存在していることが明らかとなった。さらに、リファレンス配列との比較により株ごとの変異を同定した。変異の分類は、100 塩基未満の変異、および配列の欠失または挿入 (Short Nucleotide Variation; SNV)、7 塩基以下の繰り返し領域 (Short Tandem Repeat; STR)、8 塩基以上の繰り返し領域 (Variable Number Tandem Repeat; VNTR)、100 塩基以上の配列の欠失または挿入 (Insertion/deletion; INS/DEL)、100 塩基以上の領域にわたる strand の逆転 (Inversion; INV)、100 塩基以上の領域にわたる異なる位置への配列の移動 (Translocation; TL)、100 塩基以上の配列が複製された領域 (Copy Number Variation; CNV) およびそれらが複合的に生じた変異である。その結果、リファレンス配列には存在しない新規の変異が多数確認できた (表; 一部抜粋)。

表 各株における同定された変異数

Strain	Biotype	Genome length [bp]	Genes	SNV		STR		VNTR		INS/DEL		INV		TL		TL+INV		CNV	
				KNOWN	NOVEL	KNOWN	NOVEL	KNOWN	NOVEL	KNOWN	NOVEL	KNOWN	NOVEL	KNOWN	NOVEL	KNOWN	NOVEL	KNOWN	NOVEL
NTHi-1	II	1,979,714	1,741	207,115 ¹	759 (29)	130,654 ¹	247(156)	73 ¹	21 (5)	1,459 ¹	3 (0)	20 ¹	0 (0)	93 ¹	1 (0)	113 ¹	1 (0)	10 ¹	0 (0)
NTHi-2	II	1,979,785	1,741	207,179 ¹	767 (33)	130,785 ¹	135 (50)	75 ¹	22 (5)	1,499 ¹	4 (0)	90 ¹	0 (0)	94 ¹	1 (0)	121 ¹	1 (0)	10 ¹	0 (0)
NTHi-3	II	1,853,416	1,731	197,180 ¹	2,063 (721)	122,626 ¹	66 (25)	77 ¹	26 (13)	1,468 ¹	2 (0)	103 ¹	0 (0)	80 ¹	0 (0)	51 ¹	1 (0)	0 ¹	13 (5)
NTHi-4	II	1,832,523	1,727	196,122 ¹	1,424 (475)	121,490 ¹	76 (21)	74 ¹	21 (14)	1,426 ¹	8 (0)	85 ¹	1 (0)	68 ¹	0 (0)	55 ¹	0 (1)	7 ¹	0 (0)
NTHi-5	II	1,805,933	1,714	191,827 ¹	150 (67)	119,815 ¹	19 (8)	84 ¹	12 (8)	1,316 ¹	1 (0)	71 ¹	0 (0)	105 ¹	0 (0)	96 ¹	1 (0)	7 ¹	0 (0)
NTHi-6	II	1,819,373	1,720	195,140 ¹	1,548 (596)	120,764 ¹	50 (11)	74 ¹	28 (18)	1,508 ¹	4 (0)	56 ¹	0 (0)	78 ¹	0 (0)	56 ¹	1 (0)	9 ¹	1 (1)
NTHi-7	II	1,886,470	1,738	200,652 ¹	1,810 (592)	124,924 ¹	58 (19)	76 ¹	29 (21)	1,545 ¹	5 (0)	125 ¹	0 (0)	98 ¹	0 (0)	76 ¹	0 (0)	1 ¹	12 (2)
NTHi-8	II	1,902,529	1,727	198,573 ¹	725 (404)	126,121 ¹	138 (46)	72 ¹	23 (15)	1,591 ¹	4 (0)	105 ¹	0 (0)	78 ¹	0 (0)	82 ¹	0 (0)	0 ¹	8 (5)
NTHi-9	III	1,816,081	1,722	192,456 ¹	1,132 (34)	120,659 ¹	82 (0)	77 ¹	19 (1)	1,354 ¹	5 (0)	120 ¹	0 (0)	69 ¹	2 (0)	63 ¹	2 (0)	0 ¹	7 (0)
NTHi-10	III	1,914,718	1,746	197,978 ¹	691 (409)	126,725 ¹	125 (35)	77 ¹	19 (9)	1,435 ¹	4 (0)	135 ¹	0 (0)	89 ¹	2 (2)	94 ¹	1 (0)	0 ¹	10 (2)
NTHi-11	III	1,810,981	1,725	190,880 ¹	141 (61)	120,176 ¹	68 (11)	70 ¹	26 (8)	1,380 ¹	1 (0)	29 ¹	0 (0)	82 ¹	1 (0)	75 ¹	2 (0)	0 ¹	6 (0)
NTHi-12	III	1,818,287	1,714	191,542 ¹	732 (382)	120,813 ¹	62 (12)	71 ¹	19 (11)	1,381 ¹	2 (0)	91 ¹	0 (0)	70 ¹	3 (0)	63 ¹	0 (0)	5 ¹	1 (1)
NTHi-13	III	1,822,106	1,734	193,459 ¹	1,890 (1,186)	120,947 ¹	75 (20)	67 ¹	22 (10)	1,346 ¹	5 (0)	131 ¹	0 (0)	73 ¹	0 (0)	61 ¹	3 (0)	0 ¹	8 (2)
NTHi-14	III	1,884,503	1,717	198,205 ¹	7,509 (5,245)	124,751 ¹	125 (79)	66 ¹	41 (35)	1,908 ¹	29 (0)	111 ¹	1 (0)	143 ¹	3 (3)	135 ¹	3 (1)	5 ¹	2 (2)
NTHi-15	III	1,911,901	1,704	198,897 ¹	2,600 (1,645)	126,387 ¹	115 (50)	64 ¹	39 (32)	1,586 ¹	12 (0)	113 ¹	0 (0)	89 ¹	0 (0)	104 ¹	1 (0)	6 ¹	3 (3)
NTHi-16	III	1,791,347	1,699	187,965 ¹	2,847 (1,705)	118,849 ¹	71 (35)	71 ¹	18 (14)	1,370 ¹	19 (0)	121 ¹	2 (0)	107 ¹	0 (0)	41 ¹	1 (0)	4 ¹	1 (1)
NTHi-17	IV	1,882,645	1,706	196,551 ¹	2,941 (1,652)	124,687 ¹	124 (50)	64 ¹	31 (21)	1,537 ¹	15 (0)	140 ¹	0 (0)	94 ¹	0 (0)	83 ¹	0 (1)	5 ¹	1 (1)
NTHi-18	V	1,953,416	1,743	201,753 ¹	1,874 (1,021)	129,756 ¹	157 (54)	83 ¹	27 (13)	1,547 ¹	8 (0)	114 ¹	0 (0)	139 ¹	0 (0)	99 ¹	1 (0)	7 ¹	0 (0)
NTHi-19	V	1,790,928	1,682	187,869 ¹	450 (136)	118,982 ¹	64 (7)	70 ¹	18 (6)	1,302 ¹	2 (0)	131 ¹	0 (0)	122 ¹	0 (0)	85 ¹	1 (0)	9 ¹	0 (0)
NTHi-20	V	1,936,684	1,760	198,884 ¹	646 (237)	128,230 ¹	140 (36)	73 ¹	18 (7)	1,477 ¹	1 (0)	78 ¹	0 (0)	137 ¹	0 (0)	107 ¹	2 (0)	7 ¹	1 (1)

KNOWN：リファレンス配列のいずれかに存在する変異数

NONEL：リファレンス配列には存在しない新規の変異数、そのうち（ ）内はその株にのみ見られた変異数

(4) 遺伝子変異と表現型との関連

同定した遺伝子変異と、侵襲性、病原性、毒性、薬剤耐性などの表現型との関連を解析し、その因子を同定することを目標としているが、現時点ではそれらの同定には至っていない。

インフルエンザ菌の生物型は、ウレアーゼ活性（UA）、オルニチンデカルボキシラーゼ活性（ODCA）、インドール産生（IP）の3種類の酵素活性の有無により ~ の8種類に分類される。UAに関連する遺伝子 *ureA-ureH*、ODCAに関連する遺伝子 *speF*、IPに関連する遺伝子 *tnaA* 周辺における変異についてそれぞれ調べたところ、各酵素活性が失われている株ではそれに対応する遺伝子配列が完全に欠失していることが分かり、これら3領域の遺伝子配列の有無の組み合わせによっても生物型が決定できることを確認した。生物型を決定する酵素活性は通常、呈色反応で判定されるため判定があいまいとなる場合もあるが、遺伝子配列はその判定の誤りを排除することができ、生物型の確定診断に活用できることが示された。

また、本研究で用いたインフルエンザ菌はすべて無莢膜株（NTHi）であるが、いずれも莢膜をコードする遺伝子 *bexB* が周辺を巻き込む形で完全に欠失していることが明らかとなった。通常、NTHiの判定はPCRによる *bexB* 遺伝子の増幅が陰性であることによるが、遺伝子配列の欠失を調べることでPCR偽陰性を排除し、NTHiの確定診断にできることが示された。

(5) 総括

ロングリードの次世代シーケンサーを使用することにより NTHi の全ゲノム配列を高精度に決定することができた。また、全ゲノム配列を比較することにより各株が持つ変異を同定することができた。現時点では、侵襲性、病原性、毒性、薬剤耐性などの表現型の原因となる遺伝子変異などは同定できていないが、今後、さらに全ゲノム配列を収集・蓄積することにより解析を進めていく予定である。

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計0件

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協力者	関根 章博 (Sekine Akihiro)		
研究協力者	石和田 稔彦 (Ishiwada Naruhiko)		
研究協力者	竹内 典子 (Takeuchi Noriko)		
研究協力者	竹村 亮 (Takemura Ryo)		
研究協力者	羽田 明 (Hata Akira)		